

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2001, 01:35:40 ; Search time 1698.78 Seconds  
(without alignments)  
4202.582 Million cell updates/sec

Title: US-09-388-090-3  
Perfect score: 1395  
Sequence: 1 gtgtcaaaaataccaata.....tcagggaagtcctccgtcaa 1395

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1118133 seqs, 2558875100 residues

Word size : 25  
Total number of hits satisfying chosen parameters: 7

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database :

GenEmbl.\*

1: gb\_ba1.\*  
2: gb\_ba2.\*  
3: gb\_om.\*  
4: gb\_ov.\*  
5: gb\_ph.\*  
6: gb\_pl1.\*  
7: gb\_pl2.\*  
8: gb\_pri.\*  
9: gb\_pr2.\*  
10: gb\_pr3.\*  
11: gb\_ro.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: em\_fun.\*  
15: em\_hum1.\*  
16: em\_hum2.\*  
17: em\_in.\*  
18: em\_om.\*  
19: em\_or.\*  
20: em\_ov.\*  
21: em\_pat.\*  
22: em\_ph.\*  
23: em\_pl.\*  
24: em\_ro.\*  
25: em\_scs.\*  
26: em\_sy.\*  
27: em\_un.\*  
28: em\_vi.\*  
29: gb\_ba3.\*  
30: gb\_in1.\*  
31: gb\_in2.\*  
32: gb\_in3.\*  
33: gb\_pl3.\*  
34: gb\_pr4.\*  
35: gb\_ba1.\*  
36: gb\_ba2.\*  
37: em\_htg1.\*  
38: em\_htg2.\*  
39: em\_htg3.\*  
40: em\_htg4.\*  
41: em\_htg5.\*  
42: em\_htg6.\*  
43: em\_htg7.\*

44: em\_htg8.\*  
45: em\_htg9.\*  
46: em\_htg10.\*  
47: em\_hum3.\*  
48: em\_hum4.\*  
49: em\_hum5.\*  
50: em\_hum6.\*  
51: gb\_pr5.\*  
52: gb\_pr6.\*  
53: gb\_pr7.\*  
54: gb\_htg1.\*  
55: gb\_htg2.\*  
56: gb\_htg3.\*  
57: gb\_htg4.\*  
58: gb\_htg5.\*  
59: gb\_htg6.\*  
60: gb\_htg7.\*  
61: gb\_htg8.\*  
62: gb\_htg9.\*  
63: gb\_htg10.\*  
64: gb\_htg11.\*  
65: gb\_htg12.\*  
66: gb\_htg13.\*  
67: gb\_htg14.\*  
68: gb\_htg15.\*  
69: gb\_htg16.\*  
70: gb\_htg17.\*  
71: gb\_htg18.\*  
72: gb\_htg19.\*  
73: gb\_htg20.\*  
74: gb\_htg21.\*  
75: gb\_htg22.\*  
76: gb\_htg23.\*  
77: gb\_sts1.\*  
78: gb\_sts2.\*  
79: gb\_vil.\*  
80: gb\_vil2.\*  
81: gb\_pat1.\*  
82: gb\_pat2.\*  
83: em\_htg0.\*  
84: gb\_htg24.\*  
85: gb\_pr8.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	175	12.5	1500	2	AX011572	AX011572 Sequence
2	175	12.5	1500	2	AX011574	AX011574 Sequence
c 3	175	12.5	311321	2	NMA322491	AL162754 Neisseria
4	156	11.2	1500	2	AX011576	AX011576 Sequence
c 5	156	11.2	10531	1	AE002409	AE002409 Neisseria
6	113	8.1	1110	2	AX011578	AX011578 Sequence
7	109	7.8	390	2	AX011580	AX011580 Sequence

#### ALIGNMENTS

RESULT 1  
AX011572  
LOCUS AX011572 1500 bp DNA  
DEFINITION Sequence 1 from Patent WO9955872.  
ACCESSION AX011572  
VERSION AX011572.1 GI:9998105  
KEYWORDS Neisseria meningitidis.  
SOURCE Neisseria meningitidis  
ORGANISM Neisseria meningitidis

BCT 18-SEP-2000

```

Db      294  GTTCTACCAATTTTAAACGCCCTCGTCCCGAATATGCCCCAAATTCGCCCAAGAAGAAGC
QY      354  agatgacggcggattgaacttcggttcggcggttcattcatcagcaaa 399
Db      354  AGATGACGGCGGATTTGAACATTCGGTTCCGGCTTCATCATCAGCAAA 399

RESULT  3
NMA322491/c
LOCUS   NMA322491 311321 bp          DNA          BCT          30-MAR-2000
DEFINITION  Neisseria meningitidis serogroup A strain Z2491 complete genome;
segment 3/7.
ACCESSION  AL162754 AL157959
VERSION    AL162754.2 GI:7379424
KEYWORDS
SOURCE
ORGANISM  Neisseria meningitidis.
          Neisseria meningitidis
          Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
          Neisseria.
REFERENCE  1 (bases 1 to 311321)
AUTHORS   Parkhill,J., Achtman,M., James,K.D., Bentley,S.D., Churcher,C.,
          Klee,S.R., Morelli,G., Basham,D., Brown,D., Chillingworth,T.,
          Davies,R.M., Davis,P., Devlin,K., Feltwell,T., Hamlin,N.,
          Holroyd,S., Jagels,K., Leather,S., Moule,S., Mungall,K.,
          Quail,M.A., Rajandream,M.A., Rutherford,K.M., Simmonds,M.,
          Skellton,J., Whitehead,S., Spratt,B.G. and Barrell,B.G.
TITLE      Complete DNA sequence of a serogroup A strain of Neisseria
          meningitidis Z2491
JOURNAL   Nature 404 (6777), 502-506 (2000)
MEDLINE   20222556
REFERENCE  2 (bases 1 to 311321)
AUTHORS   Parkhill,J.
TITLE      Direct Submission
COMMENT    Submitted (30-MAR-2000) Submitted on behalf of the Neisseria
          sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
          Hinxton, Cambridge CB10 1SA E-mail: parkhill@sanger.ac.uk

Notes:
Details of N. meningitidis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/N_meningitidis/).

FEATURES             source
     source            Location/Qualifiers
         1..311321
             /organism="Neisseria meningitidis"
             /strain="Z2491"
             /db_xref="taxon:487"
             /note="serogroup: A"
         115..120
             /note="serogroup: A"
         129..587
             /gene="NMA0698"
             /gene="NMA0698"
             /note="NMA0698, possible periplasmic protein, len: 152 aa;
             unknown, contains a probable N-terminal signal sequence
             and lies within a region of unusually low GC content"
         /codon_start=1
         /transl_table=11
         /product="putative periplasmic protein"
         /protein_id="CAB83985.1"
         /db_xref="GI:7379425"
         /translation="MLTKLIKLLFLFLFVFVIAINLFFFFSSDIESFGNYOPEYVY
KWGPANYILVMKDGNGFNFDKTLISGLVLEYYKEDDNIYFSYINGOCFASDSCYKPEI
LYGKILNKHNLIININSMEKNFLSEDKTMKGTRNLWADPKNKCNTQTLD"
         742..1077
             /gene="NMA0699"
             /gene="NMA0699"
             /note="NMA0699"
             /note="NMA0699, pseudogene, probable ABC transporter
             protein, len: 336 bp; similar to C-termini of many ABC
             transporters e.g. TR:Q46973 (EMBL:U47048), mIFB,
             Escherichia coli microcin transport protein (707 aa),
             fasta scores: E(): 2.6e-19, 53.2% identity in 111 aa
             overlap. Contains PS00211 ABC transporters family"

     gene              742..1077
     CDS               742..1077

```

```

signature. NMA0686 may be the remainder of this
pseudogene"
/codon_start=1
/pseudo
/transl_table=11
/product="putative ABC transporter protein (pseudogene)"
868..912
/gene="NMA0699"
/note="FS00211 ABC transporters family signature"
complement(1101..2327)
/gene="NMA0700"
complement(1101..2327)
/gene="NMA0700"
/note="NMA0700, possible ribonuclease BN, len: 408 aa;
similar to SW:RBN_ECOLI (EMBL:L19201), rbn, Escherichia
coli ribonuclease BN (EC 3.1.-.-), fasta scores; E():
2.3e-28, 34.3% identity in 268 aa overlap. Longer than rbn
at the C-terminus. Also similar to TR:085449
(EMBL:AF067083) Vitreoscilla sp. hypothetical protein (376
aa), fasta scores; E(): 0, 45.0% identity in 333 aa
overlap. Contains a region similar to NMA0677, fasta
scores; E(): 4.3e-08, 64.4% identity in 45 aa overlap.
Contains hydrophobic, possible membrane-spanning regions,
as does rbn"
/codon_start=1
/transl_table=11
/product="putative ribonuclease BN"
/protein_id="CAB83987.1"
/db_xref="GI:7379426"
/translation="MTFLQRLQGLADNKKICAFWFVVRFRDEVRPQAAASMTFTLL
ALVPLVTVMVAASIFPVFRDSDSFVSNQTVLPQGDWVFDYINAFREQANRLTA
IGSVMLVTSMLIRITDNTFNKRWNSQRMWQFLVYTWALLTFGLPSLGVGISFM
VGSVDALASGAPQWSGALRTAATLTFTLLGLYRFPVRFVPAQAFVGLALATA
FCLETARSLFTWYMGFDGYRSYGAFAVPPFLWLLNLTLVGLGAVLTSSLSYQW
GFAFRGDSRGRDDVLKILLDDAAQEGKALPVQEFRRHNMGVDELLEKLA
RHCYLYSGRGWGLTKGADSLIELNELFKLFVYRPLPVRDHVQAVDAVMPCLOTLN
MTLAEFDAQAKKQQQS"
1376..1385
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
complement(2417..2426)
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
2444..2453
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
2486..2743
/gene="NMA0701"
2486..2743
/gene="NMA0701"
/note="NMA0701, possible pseudogene, len: 258 bp; shows
weak similarity to part of SW:WRBA_ECOLI (EMBL:M99166),
wrba, Escherichia coli Trp repressor binding protein (197
aa), fasta scores; E(): 0.064, 29.3% identity in 82 aa
overlap. Also similar to part of TR:085450 (EMBL:AF067083)
Vitreoscilla sp. Trp repressor binding protein (fragment)
(124 aa), fasta scores; E(): 6.6e-14, 56.6% identity in 83
aa overlap"
/codon_start=1
/pseudo
/transl_table=11
/product="hypothetical protein NMA0701 (pseudogene)"
complement(2755..2764)
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
2778..2787
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
2861..2864
RBS
2871..3530
/gene="NMA0702"
2871..3530
/gene="NMA0702"

```

```

/note="NMA0702, len: 219 aa; similar to TR:Q52925
(EMBL:Z50189), exsB, Rhizobium melliloti putative regulator
of succinoglycan biosynthesis (not a transcriptional
regulator) (234 aa), fasta scores; E(): 3.1e-15, 32.7%
identity in 205 aa overlap. Also similar to many
bacterial ypothetical proteins e.g. SW:YBAX_HAEIN
(EMBL:U32798), H1191, Haemophilus influenzae hypothetical
protein (196 aa), fasta scores; E(): 0, 79.2% identity in
173 aa overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0702"
/protein_id="CAB83989.1"
/db_xref="GI:7379427"
/translation="MSNQKALVIFSGGDSTTCLIQAIQTYGRENVOAITFYQGRHIA
VELERAWIADGLGKQTVLDLSLRQITHNALMDDTAAITAENGVPNTFYVGRNAL
FLLYAAIYAGQGIIRIIAGVCETDFSGYPDCRDVFKSMNTVNLNAMYDFDIIHTPL
MYLTKAQTWALADENGALDIYREQTHTCYNGVGGRECPCILRERGLAEYLESKKA
V"
3522..3531
/misc_feature
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
3547..3550
RBS
3558..4076
/gene="NMA0703"
3558..4076
/gene="NMA0703"
/note="NMA0703, len: 172 aa; unknown"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0703"
/protein_id="CAB83990.1"
/db_xref="GI:7379428"
/translation="MPKLHMYLGGNAGRSNIEVHDIOFAVCDYREAYVPALKAAMFG
DADKTHDGQVVEWADGYDIASVETPKTKMSENAPRLYFANVGCYRAGOLAEAHAF
GLEAATPAEKAKQALQLLTDTSYVQOQKDNLDKVDNLLALDRIGNFHLRLTPNPHK
PAEIGFQGLPI"
3782..3791
/misc_feature
/note="NMA0703"
/note="Core DNA uptake sequence: gccgtctgaa"
/label=DUS
4083..4502
/misc_feature
/note="NMA0704"
/note="Pfam match to entry PF01242 PTPS, 6-pyruvoyl
tetrahydropterin synthase, score 10.80, E-value 1.7e-06"
4083..4505
/gene="NMA0704"
4083..4505
/gene="NMA0704"
/note="NMA0704, len: 140 aa; similar to many hypothetical
proteins e.g. SW:YB90_HAEIN (EMBL:U32798), H1190,
Haemophilus influenzae hypothetical protein (141 aa),
fasta scores; E(): 5.1e-32, 58.0% identity in 138 aa
overlap. Shows very weak similarity to eukaryotic
6-pyruvoyl-tetrahydropterin synthases e.g. SW:PTPS_RAT
(EMBL:M77850), pts, Rattus norvegicus 6-pyruvoyl
tetrahydropterin synthase precursor (EC 4.6.1.10) (144
aa), fasta scores; E(): 0.0062, 27.3% identity in 154 aa
overlap. Contains Pfam match to entry PF01242 PTPS,
6-pyruvoyl tetrahydropterin synthase"
/codon_start=1
/transl_table=11
/product="hypothetical protein NMA0704"
/protein_id="CAB83991.1"

```

```

Query Match      12.5%; Score 175; DB 2; Length 311321;
Best Local Similarity 99.6%; Pred. No. 2.6e-84;
Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 174 cdaactggttcaaacggaagcccgagtcgtcattcaggaagccccccgcgcg 233
|||||
Db 10648 CCAACTGGTTCAAGCGAAGCGCCGCGAGTCGTCAATATTCAGGAGCCCCCGCCGCG 10589

```





```
AX011578
LOCUS AX011578 1110 bp DNA BCT 18-SEP-2000
DEFINITION Sequence 7 from Patent WO955872.
ACCESSION AX011578
VERSION AX011578.1 GI:9998108
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 1110)
AUTHORS Ruelle,J.L.
TITLE Basb013 dna and proteins from neisseria meningitidis
JOURNAL Patent: WO 955872-A 04-NOV-1999;
RUELLE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
Location/Qualifiers
1..1110
source
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 280 a 350 c 279 g 201 t
ORIGIN

Query Match 8.1%; Score 113; DB 2; Length 1110;
Best Local Similarity 100.0%; Pred. NO. 2.8e-50;
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 439 atggcgagatcaaaagtccctgctcaagcagacgacgcaataataccgccaaactcgcgt 498
|||||
Db 439 ATGGCGAGTATCAAGTCCCTGCTCAAGCACAAAGCGCGAATATACCGCCAACTCATCGGT 498
|||||

QY 499 tcggatgtccaatccgatgtcgccctcttgaaaatcgacgcaacggaagact 551
|||||
Db 499 TCGGATGTCCAATCCGATGTGCGCCTTCTGAAAATCGACGCAACGGAAGACT 551
|||||

RESULT 7
AX011580
LOCUS AX011580 390 bp DNA BCT 18-SEP-2000
DEFINITION Sequence 9 from Patent WO955872.
ACCESSION AX011580
VERSION AX011580.1 GI:9998109
KEYWORDS
SOURCE
ORGANISM
Neisseria meningitidis.
Neisseria meningitidis
Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;
Neisseria.
REFERENCE
1 (bases 1 to 390)
AUTHORS Ruelle,J.L.
TITLE Basb013 dna and proteins from neisseria meningitidis
JOURNAL Patent: WO 955872-A 04-NOV-1999;
RUELLE JEAN LOUIS (BE); SMITHKLINE BEECHAM BIOLOG (BE)
FEATURES
Location/Qualifiers
1..390
source
/organism="Neisseria meningitidis"
/db_xref="taxon:487"
BASE COUNT 106 a 120 c 102 g 62 t
ORIGIN

Query Match 7.8%; Score 109; DB 2; Length 390;
Best Local Similarity 100.0%; Pred. NO. 4.5e-48;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1191 cggcgatcatcctcaaaacagatgaagccctctacacgacgaacatccggtacgtctc 1250
|||||
Db 81 CGGCGATCATCTCAAAACACATGAAGCCCCCTACACGAACGAATCCGGTACGTTCTC 140
|||||

QY 1251 ggtcgaatccgcaggcattacccttcagacacataccgacagcagcgcc 1299
|||||
Db 141 GGTGGAATCCGAGGCATTACCCCTTCAGACACATACCGACAGCAGCGGC 189
|||||
```

Search completed: March 22, 2001, 02:51:11  
Job time: 4531 sec







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2001, 01:37:50 ; Search time 90.77 Seconds  
(without alignments)  
5773.377 Million cell updates/sec

Title: US-09-388-090-3  
Perfect score: 1395  
Sequence: 1 gttgtcaaaaataccaata.....tcagggaagtcctccgtcaa 1395

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 480022 seqs, 187831343 residues

Word size : 25

Total number of hits satisfying chosen parameters: 18

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : N\_Geneseq\_36.\*  
1: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1980.DAT.\*  
2: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1981.DAT.\*  
3: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1982.DAT.\*  
4: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1983.DAT.\*  
5: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1984.DAT.\*  
6: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1985.DAT.\*  
7: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1986.DAT.\*  
8: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1987.DAT.\*  
9: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1988.DAT.\*  
10: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1989.DAT.\*  
11: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1990.DAT.\*  
12: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1991.DAT.\*  
13: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1992.DAT.\*  
14: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1993.DAT.\*  
15: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1994.DAT.\*  
16: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1995.DAT.\*  
17: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1996.DAT.\*  
18: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1997.DAT.\*  
19: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1998.DAT.\*  
20: /cgn2\_2/gcgdata/geneseq/geneseqn/NA1999.DAT.\*  
21: /cgn2\_2/gcgdata/geneseq/geneseqn/NA2000.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1395	100.0	1395	21 293414	NGSP polypeptide c
2	1344	96.3	1395	21 251538	Neisseria meningit
3	1323	94.8	1326	21 251539	Neisseria meningit
4	1248	89.5	1500	21 254510	Neisseria gonorrhe
5	1242	89.0	1242	21 293415	NGSP polypeptide c
6	175	12.5	1347	21 251533	Neisseria meningit
7	175	12.5	1500	21 233305	Neisseria meningit
8	175	12.5	1500	21 233306	Neisseria meningit
9	156	11.2	1500	21 254511	Neisseria meningit
10	156	11.2	1500	21 233307	Neisseria meningit
11	153	11.0	153	21 293416	NGSP coding sequen
12	138	9.9	1500	21 254512	Neisseria meningit

13	113	8.1	1110	21	233308	Neisseria meningit
14	109	7.8	390	21	233309	Neisseria meningit
15	38	2.7	153	21	251542	Neisseria meningit
16	33	2.4	69	21	293413	Primer used for am
17	30	2.2	30	21	251537	3'PCR primer to am
18	30	2.2	46	21	293412	Primer used for am

ALIGNMENTS

RESULT 1  
Z93414  
ID Z93414 standard; DNA; 1395 BP.  
XX  
AC Z93414;  
XX  
DT 24-JUL-2000 (first entry)  
DE  
DE NGSP polypeptide coding sequence.  
XX  
KW NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
KW screening; probe; primer; prophylaxis; therapy; ss.  
XX  
OS Neisseria gonorrhoeae.  
XX  
FH Key Location/Qualifiers  
FT 1..1395  
FT /\*tag= a  
FT /product= NGSP polypeptide  
WO200012133-A1.  
09-MAR-2000.  
01-SEP-1999; 99WO-US20070.  
01-SEP-1998; 98US-0098685.  
(ANTE-) ANTEX BIOLOGICS INC.  
Jackson WJ, Harris AM;  
WPI; 2000-237782/20.  
P-PSDB; Y83150.  
Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
Neisseria useful for diagnosis, prevention or treatment of Neisseria  
infections  
Claim 38; Page 60-61; 68pp; English.  
The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
Arg-Gly-Asn motifs near the C-terminus which function as adherence  
domains for extracellular matrix proteins. Using the NGSP polypeptide  
as a vaccine produces antibodies which inhibit binding of N.  
gonorrhoeae to the host's cellular matrix reducing attachment and/or  
subsequent invasion. The NGSP polypeptide and its peptide fragments  
can be used to immunise an animal and produce an immune response.  
They can also be used as ligands to detect antibodies elicited in  
response to Neisseria infections and also as antigens or immunogens  
for inducing Neisseria-specific antibodies which are useful in  
immunoassays to detect Neisseria in biological specimens. Nucleotides  
encoding NGSP or its fragments can be used as probes to identify  
Neisseria in biological specimens by hybridization or polymerase  
chain reaction amplification. The NGSP polypeptide can also be used  
in screening assays to identify agents and compounds which useful as  
diagnostic, prophylactic or therapeutic agents against Neisseria  
infection.  
Sequence 1395 BP; 359 A; 438 C; 353 G; 245 T; 0 other;

Query Match	100.0%	Score 1395	DB 21	Length 1395	
Best Local Similarity	100.0%	Prod. No. 0			
Matches 1395	Conservative	0	Mismatches 0	Indels 0	Gaps 0
Qy 1	gtgttcaaaaaataccaatacttcgttcttgccggcaactgtgtgcgccttctgtgcagcgc 60				
Db 1	gtgttcaaaaaataccaatacttcgttcttgccggcaactgtgtgcgccttctgtgcagcgc 60				
Qy 61	tgcgaaaaaggcaggcagcttttttcggtgcgcgacaaaaaagaagcatccttctgtagaacgc 120				
Db 61	tgcgaaaaaggcaggcagcttttttcggtgcgcgacaaaaaagaagcatccttctgtagaacgc 120				
Qy 121	atcgaaacacacacaagacgacgagctgcagtgatctgtctgcgcgaatttgcgccaatcg 180				
Db 121	atcgaaacacacacaagacgacgagctgcagtgatctgtctgcgcgaatttgcgccaatcg 180				
Qy 181	gttcaaaagcgaagcccgccgagtcgtaattcaggcagccccgcgcgcgcgcgcacccaa 240				
Db 181	gttcaaaagcgaagcccgccgagtcgtaattcaggcagccccgcgcgcgcgcgcacccaa 240				
Qy 241	aacggcagcggaatgcgcgaacccgattccgaccgcgttcgcgacgacgacccgtttctac 300				
Db 241	aacggcagcggaatgcgcgaacccgattccgaccgcgttcgcgacgacgacccgtttctac 300				
Qy 301	gaattttcaaaacgctcgtcccgaaacatgccggaatcccccaagaagaagcagatgac 360				
Db 301	gaattttcaaaacgctcgtcccgaaacatgccggaatcccccaagaagaagcagatgac 360				
Qy 361	ggcggtatgaacttcggttcgggttcctcatcagcaaaaacggctacatcctgaccaat 420				
Db 361	ggcggtatgaacttcggttcgggttcctcatcagcaaaaacggctacatcctgaccaat 420				
Qy 421	accacgtctgttcggcgtatgggcagtatcaaaagtctctgctcaacgacaaagcggaat 480				
Db 421	accacgtctgttcggcgtatgggcagtatcaaaagtctctgctcaacgacaaagcggaat 480				
Qy 481	accgccaactcatcgtcttgatgcgaatccaatccgatgtcgcccttctgaaaatcgacga 540				
Db 481	accgccaactcatcgtcttgatgcgaatccaatccgatgtcgcccttctgaaaatcgacga 540				
Qy 541	acggaagagctaccgctcgtcaaaatcggaatcccaaaatttgaaacccggcggaatgg 600				
Db 541	acggaagagctaccgctcgtcaaaatcggaatcccaaaatttgaaacccggcggaatgg 600				
Qy 601	gtcgtgcacatcgccgccttcggcttttgacaaacgctgcacgcgcgcatcgtgtcc 660				
Db 601	gtcgtgcacatcgccgccttcggcttttgacaaacgctgcacgcgcgcatcgtgtcc 660				
Qy 661	gccaaaggcagaagctctcccaacgaaagctacacaccttcataccaaacgcagcttgc 720				
Db 661	gccaaaggcagaagctctcccaacgaaagctacacaccttcataccaaacgcagcttgc 720				
Qy 721	atcaatccgggcaattccgcggcgccgctgttcaacttaaaaggacaggtcgtcggcatc 780				
Db 721	atcaatccgggcaattccgcggcgccgctgttcaacttaaaaggacaggtcgtcggcatc 780				
Qy 781	aattcgcaaatatacagccgacgagcgcggaattcatggtcatctccttgcacatcccgatt 840				
Db 781	aattcgcaaatatacagccgacgagcgcggaattcatggtcatctccttgcacatcccgatt 840				
Qy 841	gacgttgcatagaattgtccgcggaacagctgaaacacccgcaaaagtccaacgcggacaa 900				
Db 841	gacgttgcatagaattgtccgcggaacagctgaaacacccgcaaaagtccaacgcggacaa 900				
Qy 901	ctggcgctgattattcagaagaagtatcctacaggtttggcacagtgctcgtctcgataaa 960				
Db 901	ctggcgctgattattcagaagaagtatcctacaggtttggcacagtgctcgtctcgataaa 960				
Qy 961	gccagcgcgcatgattgcgcaaatcctcccgacaccccgacgaacgtgcgcgctgtg 1020				
Db 961	gccagcgcgcatgattgcgcaaatcctcccgacaccccgacgaacgtgcgcgctgtg 1020				







QY 1136 gcaagcggaagaatcacatcaagccaagctgggccaagccgagcgtaccgagc 1195  
 Db 1136 gcaagcggaagaatcacatcaagccaagctgggccaagccgagcgtaccgagc 1195  
 QY 1196 catcatcaaaacagatgaagccctcacacacgagcaacatccggtctcggtcg 1255  
 Db 1196 catcatcaaaacagatgaagccctcacacacgagcaacatccggtctcggtcg 1255  
 QY 1256 aatccgagcgattacccttcagacacatccgagcagcggaacacccctcgctcg 1315  
 Db 1256 aatccgagcgattacccttcagacacatccgagcagcggaacacccctcgctcg 1315  
 QY 1316 tacgggttccgagcgagcaagcgaggtcttaaggc 1354  
 Db 1316 tacgggttccgagcgagcaagcgaggtcttaaggc 1354

RESULT 5  
 ID 293415 standard; DNA; 1242 BP.  
 AC 293415;  
 XX  
 DT 24-JUL-2000 (first entry)  
 XX  
 NGSP polypeptide coding sequence.  
 XX  
 KW NGSP: polypeptide; peptide; vaccine; immune response; antibody;  
 KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
 KW screening; probe; primer; prophylaxis; therapy; ss.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 FH Location/Qualifiers  
 FT CDS  
 FT /\*tag= a  
 FT /product= NGSP polypeptide  
 XX  
 WO200012133-A1.  
 XX  
 PD 09-MAR-2000.  
 XX  
 PF 01-SEP-1999; 99WO-US20070.  
 XX  
 PR 01-SEP-1998; 98US-0098685.  
 XX  
 PA (ANTE-) ANTEX BIOLOGICS INC.  
 XX  
 PI Jackson WJ, Harris AM;  
 XX  
 DR P-PSDB; Y83151.  
 DR  
 XX  
 PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
 PT Neisseria useful for diagnosis, prevention or treatment of Neisseria  
 PT Infections  
 XX  
 XX  
 PS Claim 38; Page 62-63; 68pp; English.  
 XX  
 CC The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
 CC Arg-Gly-Asn motifs near the C-terminus which function as adherence  
 CC domains for extracellular matrix proteins. Using the NGSP polypeptide  
 CC as a vaccine produces antibodies which inhibit binding of N.  
 CC gonorrhoeae to the host's cellular matrix reducing attachment and/or  
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments  
 CC can be used to immunise an animal and produce an immune response.  
 CC They can also be used as ligands to detect antibodies elicited in  
 CC response to Neisseria infections and also as antigens or immunogens  
 CC for inducing Neisseria-specific antibodies which are useful in  
 CC immunoassays to detect Neisseria in biological specimens. Nucleotides  
 CC encoding NGSP or its fragments can be used as probes to identify  
 CC Neisseria in biological specimens by hybridization or polymerase

CC chain reaction amplification. The NGSP polypeptide can also be used  
 CC in screening assays to identify agents and compounds which useful as  
 CC diagnostic, prophylactic or therapeutic agents against Neisseria  
 CC Infection.  
 XX  
 SQ Sequence 1242 BP; 316 A; 400 C; 312 G; 214 T; 0 other;  
 Query Match 89.0%; Score 1242; DB 21; Length 1242;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 154 atgctgtgcccgaactttgcccaactgtttcaaaagcgaagcccgagtcgaattt 213  
 Db 1 atgctgtgcccgaactttgcccaactgtttcaaaagcgaagcccgagtcgaattt 60  
 QY 214 caggcagcccccccgccgacaccccaaaacggcagcgcaatgcgaaaccgallccgac 273  
 Db 61 caggcagcccccccgccgacaccccaaaacggcagcgcaatgcgaaaccgallccgac 120  
 QY 274 ccgcttgcgacagcgaccctgttctacgaatttttcaaacgctcgtlcccgaaacgccc 333  
 Db 121 ccgcttgcgacagcgaccctgttctacgaatttttcaaacgctcgtlcccgaaacgccc 180  
 QY 334 gaaatcccccaagaagaagcagatgacggcggttgacttcggttcggggttcacatc 393  
 Db 181 gaaatcccccaagaagaagcagatgacggcggttgacttcggttcggggttcacatc 240  
 QY 394 agcaaaaaagcgtacatccttgacaaatacccaacgctcgttccggtatggcgaatataaaa 453  
 Db 241 agcaaaaaagcgtacatccttgacaaatacccaacgctcgttccggtatggcgaatataaaa 300  
 QY 454 gtctgtcaaacgacagcgcggaataataccgccaacaaactcatcgttgcggtatgccaatccc 513  
 Db 301 gtctgtcaaacgacagcgcggaataataccgccaacaaactcatcgttgcggtatgccaatccc 360  
 QY 514 gatgtcgcccttctgaaatcgcgaacggaagagctaccctgcgtcaaaatcggcaat 573  
 Db 361 gatgtcgcccttctgaaatcgcgaacggaagagctaccctgcgtcaaaatcggcaat 420  
 QY 574 cccaaaaatttgaacccggcggaatgggtgcgtgcctcgaatccggcgcccttcggatlgac 633  
 Db 421 cccaaaaatttgaacccggcggaatgggtgcgtgcctcgaatccggcgcccttcggatlgac 480  
 QY 634 aacagcgtgaccccgccatcgttgcgcgcaaaagcgaagccttgcgaacgaagctac 693  
 Db 481 aacagcgtgaccccgccatcgttgcgcgcaaaagcgaagccttgcgaacgaagctac 540  
 QY 694 acaccccttcataccaaacccgacgttgcgcataatccggggcaatccggcgccctgtgttc 753  
 Db 541 acaccccttcataccaaacccgacgttgcgcataatccggggcaatccggcgccctgtgttc 600  
 QY 754 aacttaaaagacaggtcgtcggcatcaattcgcaaaataacagccgagcgcggttc 813  
 Db 601 aacttaaaagacaggtcgtcggcatcaattcgcaaaataacagccgagcgcggttc 660  
 QY 814 atgggcatctccttgcctccatcccgattgacgttgcgcataatgcgcgaacagctgaaa 873  
 Db 661 atgggcatctccttgcctccatcccgattgacgttgcgcataatgcgcgaacagctgaaa 720  
 QY 874 aacaccggcaaaagtcctcaaacgcggaactggcggtgattattcagggaagtatcctacggt 933  
 Db 721 aacaccggcaaaagtcctcaaacgcggaactggcggtgattattcagggaagtatcctacggt 780  
 QY 934 ttggcagctcgttccgtctlgataaaagccagcgcgcaattgattgcgcaaaatccttccc 993  
 Db 781 ttggcagctcgttccgtctlgataaaagccagcgcgcaattgattgcgcaaaatccttccc 840  
 QY 994 ggcagccccgcagaaagtcgcgcgcttgcagcgcgggcgacatcgttccctcagcctcgacggc 1053  
 Db 841 ggcagccccgcagaaagtcgcgcgcttgcagcgcgggcgacatcgttccctcagcctcgacggc 900  
 QY 1054 ggagaataactgttcttcccgcgacaccttccctccgtcatgtgctcgcgccattacgcccggaaaa 1113

Db	901	ggagaaatcgttctccgcgacacctcccgcatggtcggcgccattacgccggaaaa	
QY	1114	gaagtccagcttcggtgatggcgcaaacgagcggaagaaatcacaatcaaaagccaaactgggc	
Db	961	gaagtccagcttcggtgatggcgcaaacgagcggaagaaatcacaatcaaaagctgggc	
QY	1174	aacgcccgcgagcataccgcgcgcataccccaaaacagatgaagccccctacaccgaaacag	
Db	1021	aacgcccgcgagcataccgcgcgcataccccaaaacagatgaagccccctacaccgaaacag	
QY	1234	caatccggtacgtttccggtcgaaatccgcaggcattacccttcagacacatatcccgacagc	
Db	1081	caatccggtacgtttccggtcgaaatccgcaggcattacccttcagacacatatcccgacagc	
QY	1294	agcggcaaacacacctcgtcgttacgcggtttccgacgcgcgagaacgcgcagcgttaagg	
Db	1141	agcggcaaacacacctcgtcgttacgcggtttccgacgcgcgagaacgcgcagcgttaagg	
QY	1354	cacggcgagaaaatccttagccgtcagggcgaagtccccgtcaa	
Db	1201	cacggcgagaaaatccttagccgtcagggcgaagtccccgtcaa	
RESULT	6		
ID	Z51533		
AC	Z51533 standard; DNA; 1347 BP.		
CC	Z51533;		
XX			
DT	03-JUL-2000 (first entry)		
XX			
DE	Neisseria meningitidis NMASP protein-l encoding DNA.		
XX			
KW	NMASP: non-cytosolic; antibacterial; antiinflammatory; cytotoxic;		
KW	anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;		
KW	Neisserial infection; meningitis; septicaemia; ds.		
XX			
OS	Neisseria meningitidis.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	1..1347	
FT		/tag= a	
FT		/product= "Neisseria meningitidis protein"	
FT		/transl_except= (pos:499..501, aa: Xaa)	
FT		/note= "Xaa is unknown"	
XX			
PN	WC200012535-A2.		
XX			
PD	09-MAR-2000.		
XX			
PF	01-SEP-1999; 99WO-US19663.		
XX			
PR	01-SEP-1998; 98US-0098685.		
XX			
PA	(ANTE-) ANTEX BIOLOGICS INC.		
XX			
PI	Jackson WJ, Harris AM;		
XX			
DR	WPI; 2000-256581/22.		
DR	P-PSDB; Y70409.		
XX			
PT	Neisseria meningitidis NMASP polypeptide, nucleotide sequences and		
PT	antibodies, useful in vaccines against infection -		
XX			
XX	Clalm 41; Page 63; 75pp; English.		
XX			
CC	The present sequence is a DNA encoding Neisseria meningitidis NMASP		
CC	protein. NMASP is a non-cytosolic protein, with antibacterial and		
CC	antiinflammatory activity. It shows sequence similarity to E. coli		
CC	DegP (HtrA) protein. NMASP proteins can be used as ligands to detect		
CC	antibodies elicited in response to N. meningitidis infections. Cytotoxic		
CC	anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA		

CC may be used for diagnosis, therapy or prophylaxis of *Neisserial*  
CC infections such as, bacterial meningitis and septicaemia.  
CC Note: The protein sequence represented in SEQ ID NO:2 of the  
CC specification is erroneous"

XX  
XX  
SQ Sequence 1347 BP; 338 A; 428 C; 344 G; 236 T; 1 other;

Query Match 12.5%; Score 175; DB 21; Length 1347;  
Best Local Similarity 99.6%; Pred. No. 1e-80;  
Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps

QY 174 ccaactgttccaaagcgaagcccgccgagtcgtccaattattcaggagcccgcccgccgcg 233  
Db 21 ccaactgttccaaagcgaagcccgccgagtcgtccaattattcaggagcccgcccgccgcg 80  
QY 234 caccaaaacggcagcggaatgcggaacccgattccgagaccgcttgcgacagcgacc 293  
Db 81 caccaaaacggcagcggaatgcggaacccgattccgagaccgcttgcgacagcgacc 140  
QY 294 gttctacgaattttcaaacgcttcgtcccgaaatgcgcgaatcccccaagaagaagc 353  
Db 141 gttctacgaattttcaaacgcttcgtcccgaaatgcgcgaatcccccaagaagaagc 200  
QY 354 agatgagcggcgattgaacttcggttcgggcttcatcatcgacaaa 399  
Db 201 agatgagcggcgattgaacttcggttcgggcttcatcatcgacaaa 246

RESULT 7  
Z33305 ID Z33305 standard; DNA; 1500 BP.  
XX AC Z33305;  
XX AC Z33305;  
DT 21-FEB-2000 (first entry)  
XX  
XX  
DE Neisseria meningitidis strain ATCC 13090 BASB013 nucleotide sequence.  
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial; ss.  
XX  
XX OS Neisseria meningitidis.  
XX OS  
XX PN WO9555872-A1.  
XX  
XX PD 04-NOV-1999.  
XX  
XX PF 20-APR-1999; 99WO-EP02765.  
XX  
XX PR 23-APR-1998; 98GB-0008734.  
XX  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Ruelle J;  
XX  
XX WPI: 2000-052809/04..  
XX P-PSDB: Y52993.  
XX  
XX Novel polynucleotides and polypeptides from *Neisseria meningitidis* used  
XX to prepare vaccines against bacterial infections -  
XX  
XX Claim 15; Page 73-74; 94pp; English.  
XX  
XX The present sequence encodes a BASB013 polypeptide isolated from  
XX *Neisseria meningitidis*. BASB013 polynucleotides and polypeptides  
XX employed as research reagents and material for the discovery of  
XX treatments and diagnostics for diseases, particularly human diseases.  
XX They can be used for diagnosis of disease, staging of disease, or  
XX determining response of an infectious organism to drugs. The  
XX polynucleotides may be used as a source for hybridisation probes, and  
XX for screening of genetic mutations. serotype, organism or strain

CC identification, identification of mutation in BASB013 sequences, and as  
 CC components of arrays which are useful for diagnostic and prognostic  
 CC purposes. The polypeptides can be used to produce antibodies. The  
 CC polypeptides can also be used in vaccine formulations, and to identify  
 CC agonists and antagonists. The polypeptides, antibodies, agonists and  
 CC antagonists (which are bacteriostatic) are used for the treatment and  
 CC prevention of diseases such as upper respiratory tract infection,  
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
 CC the development and screening of antibacterial drugs. They are also used  
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
 CC thus prevent tissue damage and/or block the normal progression of  
 CC pathogenesis in infections initiated other than by the implantation of  
 CC in-dwelling devices or by other surgical techniques.  
 XX  
 SQ Sequence 1500 BP; 370 A; 468 C; 393 G; 269 T; 0 other;

Query Match 12.5%; Score 175; DB 21; Length 1500;  
 Best Local Similarity 99.6%; Pred. No. 1e-80;  
 Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 174 ccaactggttcaaaagcgaagccgagtcgtcaattcaggcagcccccgcgcg 233  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 174 ccaactggttcaaaagcgaagccgagtcgtcaattcaggcagcccccgcgcg 233  
 QY 234 caccacaaacgagcgaatgccgaacgattccgaccgcttgcgacagcagccc 293  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 234 caccacaaacgagcgaatgccgaacgattccgaccgcttgcgacagcagccc 293  
 QY 294 gttctacgaattttcacaacgctcgtccgaaacatgccgaatcccccaagaagaagc 353  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 294 gttctacgaattttcacaacgctcgtccgaaatgccgaatcccccaagaagaagc 353  
 QY 354 agatgacgagcggattgaactcgttcggttcggttcattcatcagcaaa 399  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 354 agatgacgagcggattgaactcgttcggttcggttcattcatcagcaaa 399

RESULT 8  
 Z33306  
 ID 233306 standard; DNA; 1500 BP.  
 XX  
 AC 233306;  
 DT  
 DT 21-FEB-2000 (first entry)  
 XX  
 DE Neisseria meningitidis strain ATCC 13090 BASB013 nucleotide sequence.  
 XX  
 KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
 KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
 KW invasive bacterial disease; antibacterial; ss.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO9955872-A1.  
 XX  
 PD 04-NOV-1999.  
 XX  
 PD 20-APR-1999; 99WO-EP02765.  
 PF  
 XX 23-APR-1998; 98GB-0008734.  
 PR  
 XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 PA  
 XX Ruelle J;  
 PI  
 XX WPI; 2000-052809/04.  
 DR  
 DR P-PSDB; Y52994.  
 XX  
 XX Novel polynucleotides and polypeptides from Neisseria meningitidis used  
 PT to prepare vaccines against bacterial infections  
 XX

PS Claim 12; Page 76; 94pp; English.  
 XX  
 CC The present sequence encodes a BASB013 polypeptide isolated from  
 CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be  
 CC employed as research reagents and material for the discovery of  
 CC treatments and diagnostics for diseases, particularly human diseases.  
 CC They can be used for diagnosis of disease, staging of disease, or  
 CC determining response of an infectious organism to drugs. The  
 CC polynucleotides may be used as a source for hybridisation probes, and  
 CC for screening of genetic mutations, serotype, organism or strain  
 CC identification, identification of mutation in BASB013 sequences, and as  
 CC components of arrays which are useful for diagnostic and prognostic  
 CC purposes. The polypeptides can be used to produce antibodies. The  
 CC polypeptides can also be used in vaccine formulations, and to identify  
 CC agonists and antagonists. The polypeptides, antibodies, agonists and  
 CC antagonists (which are bacteriostatic) are used for the treatment and  
 CC prevention of diseases such as upper respiratory tract infection,  
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
 CC the development and screening of antibacterial drugs. They are also used  
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
 CC thus prevent tissue damage and/or block the normal progression of  
 CC pathogenesis in infections initiated other than by the implantation of  
 CC in-dwelling devices or by other surgical techniques.  
 XX  
 SQ Sequence 1500 BP; 370 A; 469 C; 392 G; 269 T; 0 other;

Query Match 12.5%; Score 175; DB 21; Length 1500;  
 Best Local Similarity 99.6%; Pred. No. 1e-80;  
 Matches 225; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 174 ccaactggttcaaaagcgaagccgagtcgtcaattcaggcagcccccgcgcg 233  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 174 ccaactggttcaaaagcgaagccgagtcgtcaattcaggcagcccccgcgcg 233  
 QY 234 caccacaaacgagcgaatgccgaacgattccgaccgcttgcgacagcagccc 293  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 234 caccacaaacgagcgaatgccgaacgattccgaccgcttgcgacagcagccc 293  
 QY 294 gttctacgaattttcacaacgctcgtccgaaacatgccgaatcccccaagaagaagc 353  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 294 gttctacgaattttcacaacgctcgtccgaaatgccgaatcccccaagaagaagc 353  
 QY 354 agatgacgagcggattgaactcgttcggttcggttcattcatcagcaaa 399  
 Db ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 354 agatgacgagcggattgaactcgttcggttcggttcattcatcagcaaa 399

RESULT 9  
 Z54511  
 ID 254511 standard; DNA; 1500 BP.  
 XX  
 AC 254511;  
 DT  
 DT 21-MAR-2000 (first entry)  
 XX  
 DE Neisseria meningitidis ORF 986 partial DNA sequence SEQ ID NO:2969.  
 XX  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
 KW antibacterial; gene therapy; ds.  
 XX  
 OS Neisseria meningitidis.  
 XX  
 PN WO9957280-A2.  
 XX  
 PD 11-NOV-1999.  
 XX  
 PD 30-APR-1999; 99WO-US09346.  
 PF  
 XX 01-MAY-1998; 98US-0083758.  
 PR  
 PR 31-JUL-1998; 98US-0094869.



```
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX (CHIR ) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX WPI; 2000-062150/05.
DR P-PSDB; Y75749.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX Claim 7; Page 1390; 1453pp; English.
XX
XX 253015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC novel Neisseria meningitis and N. gonorrhoeae polynucleotides and
CC polypeptides Z54537 to Z54576 and Z54616 to Z5473 represent PCR
CC primers used in the exemplification of the present invention. The
CC polypeptides, the polynucleotides, antibodies and compositions of
CC the invention can be used as vaccines, as diagnostic reagents, and as
CC immunogenic compositions. The polypeptides can be used in the
CC manufacture of medicaments for treating or preventing infection due to
CC Neisserial bacteria (e.g. meningitis and septicemia), to detect the
CC presence of Neisseria bacteria, or to raise antibodies. They may also
CC be used to screen for agonists or antagonists, which may themselves
CC have use as antibacterial agents. The polynucleotides of the invention
CC may also be used in gene therapy protocols.
XX
XX Sequence 1500 BP; 386 A; 469 C; 382 G; 263 T; 0 other;
SQ
Query Match 11.2%; Score 156; DB 21; Length 1500;
Best Local Similarity 99.0%; Pred. No. 6.9e-71;
Matches 306; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 991 cccggagaccccgagaaacgtgccggcctgcaggcgaggcgagacatcgctcagcctcgac 1050
Db 991 cccggagaccccgagaaacgtgccggcctgcaggcgaggcgagacatcgctcagcctcgac 1050
Qy 1051 ggcggagaaatacgttcttcggcgacacctccgctcatggtcggcgccattacgccggaa 1110
Db 1051 ggcggagaaatacgttcttcggcgacacctccgctcatggtcggcgccattacgccggaa 1110
Qy 1111 aaagaagtcagcctgcgctatggcgcaaggcgcaagaatacacaaatcaaaagccaagctg 1170
Db 1111 aaagaagtcagcctgcgctatggcgcaaggcgcaagaatacacaaatcaaaagccaagctg 1170
Qy 1171 ggcgaacgcgcgcagacataccggcgacatacccaaaacagatgaagccocctacaccgaa 1230
Db 1171 ggcgaacgcgcgcagacataccggcgacatacccaaaacagatgaagccocctacaccgaa 1230
Qy 1231 cagcaaatccggtacgttcttcggtcgaatcccgagcattacccttcagacacataccgac 1290
Db 1231 cagcaaatccggtacgttcttcggtcgaatcccgagcattacccttcagacacataccgac 1290
Qy 1291 agcagcggc 1299
Db 1291 agcagcggc 1299
RESULT 10
ID Z33307 standard; DNA; 1500 BP.
XX
AC Z33307;
```

```
XX 21-FEB-2000 (first entry)
XX
XX Neisseria meningitidis strain H44/76 BASB013 nucleotide sequence.
DE
DE Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
KW invasive bacterial disease; antibacterial; ss.
XX
XX Neisseria meningitidis.
OS
XX WO9955872-A1.
PN
XX 04-NOV-1999.
PD
XX 20-APR-1999; 99WO-EP02765.
PF
XX 23-APR-1998; 98GB-0008734.
PR
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
PA
XX Ruelle J;
PI
XX WPI: 2000-C52809/04.
DR P-PSDB; Y52995.
XX
XX Novel polynucleotides and polypeptides from Neisseria meningitis used
PT to prepare vaccines against bacterial infections -
XX
XX Claim 12; Page 78-79; 94pp; English.
XX
XX The present sequence encodes a BASB013 polypeptide isolated from
CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be
CC employed as research reagents and material for the discovery of
CC treatments and diagnostics for diseases, particularly human diseases.
CC They can be used for diagnosis of disease, staging of disease, or
CC determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and
CC for screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB013 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies. The
CC polypeptides can also be used in vaccine formulations, and to identify
CC agonists and antagonists. The polypeptides, antibodies, agonists and
CC antagonists (which are bacteriostatic) are used for the treatment and
CC prevention of diseases such as upper respiratory tract infection,
CC invasive bacterial diseases such as bacteraemia and meningitis, and for
CC the development and screening of antibacterial drugs. They are also used
CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
CC on in-dwelling devices, or to extracellular proteins on wounds, and to
CC thus prevent tissue damage and/or block the normal progression of
CC pathogenesis in infections initiated other than by the implantation of
CC in-dwelling devices or by other surgical techniques.
XX
XX Sequence 1500 BP; 385 A; 471 C; 382 G; 262 T; 0 other;
SQ
Query Match 11.2%; Score 156; DB 21; Length 1500;
Best Local Similarity 99.0%; Pred. No. 6.9e-71;
Matches 306; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 991 cccggagaccccgagaaacgtgccggcctgcaggcgaggcgagacatcgctcagcctcgac 1050
Db 991 cccggagaccccgagaaacgtgccggcctgcaggcgaggcgagacatcgctcagcctcgac 1050
Qy 1051 ggcggagaaatacgttcttcggcgagacacctcccgctcatggtcggcgccattacgccggaa 1110
Db 1051 ggcggagaaatacgttcttcggcgagacacctcccgctcatggtcggcgccattacgccggaa 1110
Qy 1111 aaagaagtcagcctgcgctatggcgcaaggcgcaagaatacacaaatcaaaagccaagctg 1170
Db 1111 aaagaagtcagcctgcgctatggcgcaaggcgcaagaatacacaaatcaaaagccaagctg 1170
```

QY 1171 ggcaagccgcgagacataccggcgcatcctcccaaaacagatgaagcccccacaccgaa 1230  
|||||  
Db 1171 ggcaagccgcgagacataccggcgcatcctcccaaaacagatgaagcccccacaccgaa 1230  
QY 1231 cagcaatccggtacgttctcgtcggaatccgcagcagcattacccttcagacacataccgac 1290  
|||||  
Db 1231 cagcaatccggtacgttctcgtcggaatccgcagcagcattacccttcagacacataccgac 1290  
QY 1291 agcagcggc 1299  
|||||  
Db 1291 agcagcggc 1299  
  
RESULT 11  
Z93416  
ID Z93416 standard; DNA; 153 BP.  
XX  
AC Z93416;  
XX  
DT 24-JUL-2000 (first entry)  
XX  
DE NGSP coding sequence fragment.  
XX  
KW NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
cellular matrix; adherence domain; ligand; detection; diagnosis;  
screening; probe; primer; prophylaxis; therapy; ss.  
XX  
OS Neisseria gonorrhoeae.  
FH Key Location/Qualifiers  
FT 1..1242  
FT /\*tag= a  
FT /product= NGSP polypeptide  
XX  
PN WO200012133-A1.  
XX  
PD 09-MAR-2000.  
XX  
PF 01-SEP-1999; 99WO-US20070.  
XX  
PR 01-SEP-1998; 98US-0098685.  
XX  
PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PI Jackson WJ, Harris AM;  
XX  
DR WPI; 2000-237782/20.  
DR P-PSDB; Y83152.  
XX  
PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
Neisseria useful for diagnosis, prevention or treatment of Neisseria  
infections  
XX  
PS Claim 38; Page 65; 68pp; English.  
XX  
CC The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
Arg-Gly-Asn motifs near the C-terminus which function as adherence  
domains for extracellular matrix proteins. Using the NGSP polypeptide  
as a vaccine produces antibodies which inhibit binding of N.  
gonorrhoeae to the host's cellular matrix reducing attachment and/or  
subsequent invasion. The NGSP polypeptide and its peptide fragments  
can be used to immunise an animal and produce an immune response.  
They can also be used as ligands to detect antibodies elicited in  
response to Neisseria infections and also as antigens or immunogens  
for inducing Neisseria-specific antibodies which are useful in  
immunoassays to detect Neisseria in biological specimens. Nucleotides  
encoding NGSP or its fragments can be used as probes to identify  
Neisseria in biological specimens by hybridization or polymerase  
chain reaction amplification. The NGSP polypeptide can also be used  
in screening assays to identify agents and compounds which useful as  
diagnostic, prophylactic or therapeutic agents against Neisseria  
infection.  
XX

SQ Sequence 153 BP; 43 A; 38 C; 41 G; 31 T; 0 other;  
  
Query Match 11.0%; Score 153; DB 21; Length 153;  
Best Local Similarity 100.0%; Pred. No. 2.3e-69;  
Matches 153; Conservative 0; Mismatches 0; Indels 0; Caps 0;  
  
QY 1 ggtttcaaaaaatacccaatacttcgttggcgccactgtgtccgcttgcctgagcaggc 60  
|||||  
Db 1 ggtttcaaaaaatacccaatacttcgttggcgccactgtgtccgcttgcctgagcaggc 60  
QY 61 tgcgaaaaggcaggcgagcgttttctcggtgcggacaaaagaagcagcagcagc 120  
|||||  
Db 61 tgcgaaaaggcaggcgagcgttttctcggtgcggacaaaagaagcagcagcagc 120  
QY 121 atcgaacacacacaaagacgacgagcagtcagtcagt 153  
|||||  
Db 121 atcgaacacacacaaagacgacgagcagtcagtcagt 153  
  
RESULT 12  
Z54512  
ID Z54512 standard; DNA; 1500 BP.  
XX  
AC Z54512;  
XX  
DT 21-MAR-2000 (first entry)  
XX  
DE Neisseria meningitidis ORF 986 partial DNA sequence SEQ ID NO:2971.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
antibacterial; gene therapy; ds.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9957280-A2.  
XX  
PD 11-NOV-1999.  
XX  
PF 30-APR-1999; 99WO-US09346.  
XX  
PR 01-MAY-1998; 98US-0083758.  
PR 31-JUL-1998; 98US-0094869.  
PR 02-SEP-1998; 98US-0098994.  
PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX  
PA (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;  
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarcellini M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI; 2000-062150/05.  
DR P-PSDB; Y75750.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
vaccines and diagnostics  
XX  
PS Claim 7; Page 1392; 1453pp; English.  
XX  
CC Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent  
novel Neisseria meningitis and N. gonorrhoeae polynucleotides and  
polypeptides. Z54537 to Z54576 and Z54616 to Z5473 represent PCR  
primers used in the exemplification of the present invention. The  
polypeptides, the polynucleotides, antibodies and compositions of  
the invention can be used as vaccines, as diagnostic reagents, and as  
immunogenic compositions. The polypeptides can be used in the

CC manufacture of medicaments for treating or preventing infection due to  
CC Neisseria bacteria (e.g. meningitis and septicaemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
SQ Sequence 1500 BP; 382 A; 469 C; 382 G; 265 T; 2 other;  
  
Query Match 9.9%; Score 138; DB 21; Length 1500;  
Best Local Similarity 99.5%; Pred. No. 1.4e-61;  
Matches 188; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 174 ccaactggttcaaaagcgaagccggcagtcgttcgaatttcagggcagcccccgcgcg 233  
|||||  
Db 174 ccaactggttcaaaagcgaagccggcagtcgttcgaatttcagggcagcccccgcgcg 233  
|||||  
Qy 234 caccacaaagcgcagcgcaatgcgaaacggattccgaccgcttgcgacagcgaccc 293  
|||||  
Db 234 caccacaaagcgcagcgcaatgcgaaacggattccgaccgcttgcgacagcgaccc 293  
|||||  
Qy 294 gttctacgaattttcaaacgcctctgcgcgaacatgcgcgaaatcccccaagaagaagc 353  
|||||  
Db 294 gttctacgaattttcaaacgcctctgcgcgaacatgcgcgaaatcccccaagaagaagc 353  
|||||  
Qy 354 agatgacgg 362  
|||||  
Db 354 agatgacgg 362  
|||||  
  
RESULT 13  
Z33308  
ID Z33308 standard; DNA; 1110 BP.  
XX  
AC Z33308;  
XX  
DT 21-FEB-2000 (first entry)  
XX  
DE Neisseria meningitidis BASB013-C nucleotide sequence.  
XX  
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial; ss.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9955872-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 20-APR-1999; 99WO-EP02765.  
XX  
PR 23-APR-1998; 98GB-0008734.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J;  
XX  
DR WPI; 2000-052809/04.  
DR P-PSDB; Y52996.  
XX  
XX Novel polynucleotides and polypeptides from Neisseria meningitis used  
PT to prepare vaccines against bacterial infections -  
XX  
PS Example 2; Page 81; 94pp; English.  
XX  
CC The present sequence encodes a conserved BASB013-C polypeptide isolated  
CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides  
CC may be employed as research reagents and material for the discovery of  
CC treatments and diagnostics for diseases, particularly human diseases.  
CC They can be used for diagnosis of disease, staging of disease, or  
CC determining response of an infectious organism to drugs. The

CC polynucleotides may be used as a source for hybridisation probes, and  
CC for screening of genetic mutations, serotype, organism or strain  
CC identification, identification of mutation in BASB013 sequences, and as  
CC components of arrays which are useful for diagnostic and prognostic  
CC purposes. The polypeptides can be used to produce antibodies. The  
CC polypeptides can also be used in vaccine formulations, and to identify  
CC agonists and antagonists. The polypeptides, antibodies, agonists and  
CC antagonists (which are bacteriostatic) are used for the treatment and  
CC prevention of diseases such as upper respiratory tract infection,  
CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
CC the development and screening of antibacterial drugs. They are also used  
CC in the prevention and adhesion of bacteria to eukaryotic matrix proteins  
CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
CC thus prevent tissue damage and/or block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or by other surgical techniques.  
XX  
SQ Sequence 1110 BP; 280 A; 350 C; 279 G; 201 T; 0 other;  
  
Query Match 8.1%; Score 113; DB 21; Length 1110;  
Best Local Similarity 100.0%; Pred. No. 1.1e-48;  
Matches 113; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 439 atgggcagtgatcaaaagtcctgctcaacgacgaagcgcgaataacgcgaactcgcgt 498  
|||||  
Db 439 atgggcagtgatcaaaagtcctgctcaacgacgaagcgcgaataacgcgaactcgcgt 498  
|||||  
Qy 499 tcggatgtccaaatccgatgcgccttctgaaatcgacgcaacggaagagct 551  
|||||  
Db 499 tcggatgtccaaatccgatgcgccttctgaaatcgacgcaacggaagagct 551  
|||||  
  
RESULT 14  
Z33309  
ID Z33309 standard; DNA; 390 BP.  
XX  
AC Z33309;  
XX  
DT 21-FEB-2000 (first entry)  
XX  
DE Neisseria meningitidis variable BASB013-V nucleotide sequence.  
XX  
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial; ss.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO9955872-A1.  
XX  
PD 04-NOV-1999.  
XX  
PF 20-APR-1999; 99WO-EP02765.  
XX  
PR 23-APR-1998; 98GB-0008734.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J;  
XX  
DR WPI; 2000-052809/04.  
DR P-PSDB; Y52997.  
XX  
XX Novel polynucleotides and polypeptides from Neisseria meningitis used  
PT to prepare vaccines against bacterial infections -  
XX  
PS Example 2; Page 83; 94pp; English.  
XX  
CC The present sequence encodes a variable BASB013-V polypeptide isolated  
CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides  
CC may be employed as research reagents and material for the discovery of  
CC treatments and diagnostics for diseases, particularly human diseases.

They can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BARS013 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies. The polypeptides can also be used in vaccine formulations, and to identify agonists and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases such as upper respiratory tract infection, invasive bacterial diseases such as bacteraemia and meningitis, and for the development and screening of antibacterial drugs. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogenesis in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

```
Query Match          7.8%; Score 109; DB 21; Length 390;
Best Local Similarity 100.0%; Pred. No. 1.3e-46;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1191	cggcgcatctccaaaacagatgaagccccctacacggaaacagcaatccggtagcttctc	1250
Db	81	cggcgcatctccaaaacagatgaagccccctacacggaaacagcaatccggtagcttctc	140
Qy	1251	ggtcgaatcccgaggcatatcccttcagacacatccgacagcagcgcc	1299
Db	141	ggtcgaatcccgaggcatatcccttcagacacatccgacagcagcgcc	189

RESULT 15

RESOLUTION 13  
251542  
ID 251542 standard; DNA; 153 BP.

AC 251542;

03-JUL-2000 (first entry)

DE *Neisseria meningitidis* NMASP protein N-terminal fragment encoding DNA.

AA NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia; ds.  
KW

OS *Neisseria meningitidis*.

PN WO200012535-A2.

09-MAR-2000.

AA  
PF 01-SEP-1999; 99WO-US19663.AA  
PR 01-SEP-1998; 98US-0098685.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson WJ, Harris AM;

WPI; 2000-256581/22.

DR F-PSDB; Y70413.  
XX

Neisseria meningitidis NMASP polypeptide, nucleotide sequences and antibodies, useful in vaccines against infection -

PS Claim 42; Page 74; 75pp; English.

CC The present sequence is a DNA encoding *Neisseria meningitidis* NMASP  
CC protein N-terminal fragment. NMASP is a non-cytosolic protein, with  
CC

antibacterial and antiinflammatory activity. It shows sequence similarity to *E. coli* Degrp (Htra) protein. MNASP proteins can be used as ligands to detect antibodies elicited in response to *N. meningitidis* infections. Cytotoxic anti-MNASP antibodies can be used as vaccines. MNASP proteins and DNA may be used for diagnosis, therapy or prophylaxis of *Neisseria* infections such as, bacterial meningitis and septicæmia.

Sequence 153 BP; 44 A; 42 C; 38 G; 29 T; 0 other;

Query Match 2.7%; Score 38; DB 21; Length 153;  
Best Local Similarity 100.0%; Pred. No. 6.4e-10;

Qy 73 ggcagcttttttcggtgcggacaaaaaagaagcatcctt 110  
|||||

Db 73 ggagcgttttcggtgcggacaaaaaaggacatcctt 110

Search completed: March 22, 2001, 02:52:06  
Job time: 4456 sec

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2001, 01:36:30 ; Search time 58.8 Seconds  
(without alignments)  
3823.446 Million cell updates/sec

Title: US-09-388-090-3  
Perfect score: 1395  
Sequence: 1 gtgttcaaaaataccaata.....tcagggcaagtcctccgtcaa 1395

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 280836 seqs, 80580151 residues

Word size : 25

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6\_COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES			
Result	Query		
No.	Score	Match Length DB	ID Description
-----			
No matches found			

Search completed: March 22, 2001, 02:50:12  
Job time: 4422 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 22, 2001, 01:22:25 ; Search time 885.72 Seconds  
(without alignments)  
11036.722 Million cell updates/sec

Title: US-09-388-090-3  
Perfect score: 1395  
Sequence: 1.gtgttcaaaaaataccaata.....tcaggccaagtccccgtcaa 1395

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 7991742 seqs, 3503743858 residues

Word size : 25

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : EST.\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_est4:\*  
5: gb\_est5:\*  
6: gb\_est6:\*  
7: gb\_est7:\*  
8: gb\_est8:\*  
9: gb\_est9:\*  
10: gb\_est10:\*  
11: gb\_est11:\*  
12: gb\_est12:\*  
13: gb\_est13:\*  
14: gb\_est14:\*  
15: gb\_est15:\*  
16: gb\_est16:\*  
17: gb\_est17:\*  
18: gb\_est18:\*  
19: gb\_est19:\*  
20: gb\_est20:\*  
21: gb\_est21:\*  
22: gb\_est22:\*  
23: gb\_est23:\*  
24: gb\_est24:\*  
25: gb\_est25:\*  
26: gb\_est26:\*  
27: gb\_est27:\*  
28: gb\_est28:\*  
29: gb\_est29:\*  
30: gb\_est30:\*  
31: gb\_est31:\*  
32: gb\_est32:\*  
33: gb\_est33:\*  
34: gb\_est34:\*  
35: gb\_est35:\*  
36: gb\_est36:\*  
37: gb\_est37:\*  
38: gb\_est38:\*  
39: gb\_est39:\*  
40: gb\_est40:\*  
41: em\_estba:\*  
42: em\_estfun:\*  
43: em\_esthum1:\*  
44: em\_esthum2:\*  
45: em\_esthum3:\*  
46: em\_esthum4:\*  
47: em\_esthum5:\*  
48: em\_esthum6:\*  
49: em\_esthum7:\*  
50: em\_esthum8:\*  
51: em\_esthum9:\*  
52: em\_esthum10:\*  
53: em\_esthum11:\*  
54: em\_esthum12:\*  
55: em\_esthum13:\*  
56: em\_esthum14:\*  
57: em\_esthum15:\*  
58: em\_esthum16:\*  
59: em\_esthum17:\*  
60: em\_esthum18:\*  
61: em\_esthum19:\*  
62: em\_esthum20:\*  
63: em\_estin1:\*  
64: em\_estin2:\*  
65: em\_estin3:\*  
66: em\_estin4:\*  
67: em\_estov1:\*  
68: em\_estov2:\*  
69: em\_estpl1:\*  
70: em\_estpl2:\*  
71: em\_estpl3:\*  
72: em\_estpl4:\*  
73: em\_estpl5:\*  
74: em\_estro1:\*  
75: em\_estro2:\*  
76: em\_estro3:\*  
77: em\_estro4:\*  
78: em\_estro5:\*  
79: em\_estro6:\*  
80: em\_estro7:\*  
81: em\_estro8:\*  
82: em\_estro9:\*  
83: em\_estro10:\*  
84: em\_estro11:\*  
85: em\_estro12:\*  
86: em\_estro13:\*  
87: gb\_est41:\*  
88: gb\_est42:\*  
89: gb\_est43:\*  
90: gb\_est44:\*  
91: gb\_est45:\*  
92: gb\_est46:\*  
93: gb\_est47:\*  
94: gb\_est48:\*  
95: gb\_est49:\*  
96: gb\_est50:\*  
97: gb\_est51:\*  
98: gb\_est52:\*  
99: gb\_est53:\*  
100: gb\_est54:\*  
101: gb\_est55:\*  
102: gb\_est56:\*  
103: gb\_est57:\*  
104: gb\_est67:\*  
105: gb\_est68:\*  
106: gb\_est69:\*  
107: gb\_est70:\*  
108: gb\_est71:\*  
109: gb\_est72:\*  
110: gb\_est73:\*  
111: gb\_est74:\*  
112: em\_esthum21:\*  
113: em\_esthum22:\*  
114: em\_esthum23:\*  
115: em\_estom1:\*  
116: em\_estom2:\*

117: em\_estpl6:\*  
 118: em\_estpl7:\*  
 119: em\_estpl8:\*  
 120: em\_estrol4:\*  
 121: em\_estrol5:\*  
 122: em\_estrol6:\*  
 123: em\_estrol7:\*  
 124: em\_estrol8:\*  
 125: em\_estrol9:\*  
 126: gb\_est58:\*  
 127: gb\_est59:\*  
 128: gb\_est60:\*  
 129: gb\_est61:\*  
 130: gb\_est62:\*  
 131: gb\_est63:\*  
 132: gb\_est64:\*  
 133: gb\_est65:\*  
 134: gb\_est66:\*  
 135: gb\_est75:\*  
 136: gb\_est76:\*  
 137: gb\_est77:\*  
 138: gb\_est78:\*  
 139: gb\_est79:\*  
 140: gb\_est80:\*  
 141: gb\_est81:\*  
 142: gb\_est82:\*  
 143: gb\_est83:\*  
 144: gb\_est84:\*  
 145: gb\_est85:\*  
 146: gb\_est86:\*  
 147: gb\_est87:\*  
 148: gb\_gss1:\*  
 149: gb\_gss2:\*  
 150: gb\_gss3:\*  
 151: gb\_gss4:\*  
 152: em\_gss1:\*  
 153: em\_gss2:\*  
 154: em\_gss3:\*  
 155: em\_gss4:\*  
 156: gb\_gss5:\*  
 157: gb\_gss6:\*  
 158: gb\_gss7:\*  
 159: gb\_gss8:\*  
 160: gb\_gss9:\*  
 161: gb\_gss10:\*  
 162: gb\_gss11:\*  
 163: gb\_gss12:\*  
 164: gb\_gss13:\*  
 165: gb\_gss14:\*  
 166: gb\_gss15:\*  
 167: gb\_gss16:\*  
 168: gb\_gss17:\*  
 169: gb\_gss18:\*  
 170: gb\_gss19:\*  
 171: gb\_gss20:\*  
 172: gb\_gss21:\*  
 173: gb\_gss22:\*  
 174: gb\_gss23:\*  
 175: gb\_gss24:\*  
 176: em\_gss5:\*  
 177: em\_gss6:\*  
 178: em\_gss7:\*  
 179: em\_gss8:\*  
 180: em\_gss9:\*  
 181: em\_gss10:\*  
 182: em\_gss11:\*  
 183: em\_gss12:\*  
 184: em\_gss13:\*  
 185: em\_gss14:\*  
 186: em\_gss15:\*  
 187: em\_gss16:\*  
 188: em\_gss17:\*  
 189: em\_gss18:\*

190: gb\_gss25:\*  
 191: gb\_gss26:\*  
 192: gb\_gss27:\*  
 193: gb\_gss28:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description
-----							-----

No matches found

Search completed: March 22, 2001, 02:20:30  
 Job time: 3485 sec

---



OM of: US-09-388-090-3 to: A\_Geneseq\_36.\* out\_format : pfs  
Date: Mar 22, 2001 12:04 PM  
About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.  
Command line parameters:  
-MODEL=frame+nt2p.model -DEV=xlp  
-O=/cgn2\_1/JUSTO\_spool-US09388090/funat\_20032001\_092950\_25784/app\_query.fasta\_1.1474  
-DB=A\_Geneseq\_36 -QFMT=fastan -SUFFIX=oligo7.rag -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=6.000 -GAPEXT=0.050 -GAPOP=60.000 -XGAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -YGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=1000 -DOCLALIGN=200 -THR\_SCORE=quality  
-THR\_MIN=7 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09388090 @CNG1\_1\_36 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-388-090-3

Query length: 1395

Database: A\_Geneseq\_36.\*

Database sequences: 268485

Database length: 34193795

Search time (sec): 33.030000

WARN: XGAPOP and XGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd Orig	Zscore	Escore Len	Documentation
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y83150 +	465.00	8134.18	0.0	465
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75748 +	451.00	7887.98	0.0	499
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y83151 +	414.00	7239.91	0.0	414
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52996 +	127.00	2203.72	1.1e-115	370
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75749 +	127.00	2201.64	1.1e-115	499
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52994 +	127.00	2201.64	1.1e-115	499
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52995 +	127.00	2201.64	1.1e-115	499
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52993 +	115.00	1991.04	5.8e-104	448
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y70409 +	103.00	1781.18	3.2e-92	499
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y70414 +	103.00	1780.77	3.1e-92	475
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y70413 +	103.00	1780.45	3.1e-92	498
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y57506 +	51.00	883.64	2.7e-41	51
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52997 +	36.00	613.94	1.1e-26	129
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y70415 +	19.00	322.02	5.2e-10	51
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y29294 +	18.00	289.21	3.9e-09	460
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y83154 +	11.00	192.26	0.0410	11
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y60397 -	9.00	132.07	2.48	128
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y91334 +	9.00	131.68	2.47	433
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y59354 +	9.00	131.68	2.47	433
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y59353 +	9.00	131.68	2.47	433
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y59355 +	9.00	131.68	2.47	433
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y04995 +	9.00	129.65	2.39	580
/cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT:Y32431 +	9.00	129.65	2.39	580
/cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT:Y32363 +	9.00	129.65	2.39	580
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y81666 +	9.00	129.65	2.39	580
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y63103 +	9.00	129.65	2.39	580
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y38968 +	9.00	129.65	2.39	580
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y04996 +	8.00	116.95	24.50	288
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y89984 +	8.00	116.95	24.50	297
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y64449 +	8.00	116.74	23.92	358
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y80102 +	8.00	115.45	23.92	357
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y98445 +	8.00	114.90	23.72	387

/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:W89850 +	8.00	113.97	23.37	443
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:W89877 +	8.00	113.84	23.32	451
/cgn2_2/gcgdata/geneseq/geneseq/AA1996.DAT:R77434 +	8.00	113.99	23.04	503
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y55721 +	8.00	112.43	22.80	553
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y41277 +	8.00	112.43	22.80	553
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y43280 +	8.00	112.43	22.80	553
/cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:R51355 +	8.00	112.39	22.79	556
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y53675 -	8.00	100.11	18.76	3262
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y39762 +	8.00	100.11	18.75	3264
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R39797 -	7.00	122.06	333.33	11
/cgn2_2/gcgdata/geneseq/geneseq/AA1991.DAT:R14324 -	7.00	117.91	312.10	21
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y20222 -	7.00	117.57	310.43	21
/cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:R63153 -	7.00	115.83	301.95	27
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R46589 +	7.00	115.58	300.75	56
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y12582 -	7.00	110.77	278.65	58
/cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:R58674 +	7.00	108.56	269.05	77
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y21088 +	7.00	107.95	266.49	84
/cgn2_2/gcgdata/geneseq/geneseq/AA1989.DAT:R90408 -	7.00	107.63	265.12	88
/cgn2_2/gcgdata/geneseq/geneseq/AA1995.DAT:R80528 -	7.00	107.17	263.21	94
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y60272 +	7.00	107.103	262.60	96
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:R20888 -	7.00	107.03	262.60	96
/cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT:R29888 -	7.00	106.54	260.57	103
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y44878 +	7.00	106.21	259.21	108
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R98318 +	7.00	105.90	257.92	113
/cgn2_2/gcgdata/geneseq/geneseq/AA1992.DAT:R26087 -	7.00	105.90	257.92	113
/cgn2_2/gcgdata/geneseq/geneseq/AA1993.DAT:R38722 -	7.00	105.90	257.92	113
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R41979 +	7.00	105.90	257.92	113
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:R76993 +	7.00	105.90	257.92	113
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R58853 +	7.00	105.77	257.43	115
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R38606 +	7.00	105.42	255.99	121
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y50973 +	7.00	105.36	255.76	122
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y50975 +	7.00	105.36	255.76	122
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75703 -	7.00	105.03	254.41	128
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:R64530 +	7.00	104.17	250.94	145
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R81599 +	7.00	103.61	248.75	157
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R56766 +	7.00	102.94	246.11	173
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y36820 -	7.00	102.44	244.15	186
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y48501 +	7.00	102.18	243.16	193
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75702 -	7.00	101.87	241.95	202
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75704 -	7.00	101.87	241.95	202
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75705 -	7.00	101.87	241.95	202
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75706 -	7.00	101.87	241.95	202
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75707 -	7.00	101.87	241.95	202
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75205 -	7.00	101.83	241.81	203
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y12916 +	7.00	101.73	241.42	206
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y12969 +	7.00	101.73	241.42	206
/cgn2_2/gcgdata/geneseq/geneseq/AA1994.DAT:R47240 -	7.00	101.50	240.54	213
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:R93648 +	7.00	101.30	239.80	219
/cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT:R06438 +	7.00	101.27	239.80	220
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R69515 +	7.00	101.21	239.68	222
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75485 +	7.00	101.21	239.44	222
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75486 +	7.00	101.21	239.44	222
/cgn2_2/gcgdata/geneseq/geneseq/AA1997.DAT:W01819 -	7.00	100.70	237.51	239
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:Y63762 -	7.00	100.70	237.51	239
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y23988 +	7.00	100.19	235.62	257
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:Y23987 +	7.00	100.11	235.31	260
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R62827 -	7.00	100.03	235.02	263
/cgn2_2/gcgdata/geneseq/geneseq/AA1999.DAT:R88481 +	7.00	99.90	234.53	268
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y80963 +	7.00	99.72	233.77	285
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y56056 -	7.00	99.43	232.77	287
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y56077 -	7.00	99.43	232.77	287
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y57853 -	7.00	99.43	232.77	287
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y57865 -	7.00	99.43	232.77	287
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:R80961 +	7.00	99.36	232.50	290
/cgn2_2/gcgdata/geneseq/geneseq/AA1990.DAT:R03340 -	7.00	99.33	232.41	291
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:R68809 -	7.00	99.17	231.81	298
/cgn2_2/gcgdata/geneseq/geneseq/AA1998.DAT:W62268 -	7.00	99.17	231.81	298
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y56057 -	7.00	99.01	231.22	305
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y57866 -	7.00	99.01	231.22	305
/cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y57867 +	7.00	98.80	230.48	314

/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y91672 +	7.00	98.80	230.48	314	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y06443 -	7.00	95.15	217.48	532
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W88749 -	7.00	98.76	230.32	316	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1997.DAT.W25769 +	7.00	95.12	217.39	534
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80962 +	7.00	98.67	230.00	320	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W39788 +	7.00	94.97	216.86	546
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56767 +	7.00	98.61	229.76	323	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W40112 -	7.00	94.95	216.81	547
/cgn2_2/cgdata/geneseq/geneseq/AA1992.DAT.R29266 +	7.00	98.46	229.22	330	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R58661 +	7.00	94.89	216.60	552
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W58564 +	7.00	98.46	229.22	330	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R58661 +	7.00	94.88	216.55	553
/cgn2_2/cgdata/geneseq/geneseq/AA1992.DAT.R26874 +	7.00	98.31	228.69	337	I	A	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R58662 +	7.00	94.88	216.55	553
/cgn2_2/cgdata/geneseq/geneseq/AA1993.DAT.R35044 +	7.00	98.31	228.69	337	I	A	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W96316 -	7.00	94.73	216.04	565
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y77184 +	7.00	98.31	228.69	337	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W96315 -	7.00	94.66	215.79	571
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80939 +	7.00	98.31	228.69	337	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W82547 +	7.00	94.33	214.7	591
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y27804 -	7.00	98.23	228.39	341	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.W35062 +	7.00	94.30	214.58	601
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W48667 -	7.00	98.19	228.25	343	I	C	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W42491 +	7.00	94.30	214.58	601
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80959 +	7.00	98.19	228.25	343	I	C	/cgn2_2/cgdata/geneseq/geneseq/AA1997.DAT.W42490 -	7.00	94.27	214.46	604
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.Y80956 -	7.00	98.17	228.17	344	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W62779 +	7.00	94.14	214.03	615
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P90506 -	7.00	97.84	226.96	361	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W62780 +	7.00	94.14	214.03	615
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P93566 -	7.00	97.82	226.89	362	I	A	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y57301 -	7.00	93.86	213.10	640
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P93567 -	7.00	97.82	226.89	362	I	A	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y57301 -	7.00	93.72	212.63	653
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P90419 -	7.00	97.82	226.89	362	I	F	/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y91524 +	7.00	93.26	211.07	698
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y40499 +	7.00	97.74	226.62	366	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1997.DAT.W23599 -	7.00	93.21	210.91	703
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y04940 +	7.00	97.61	226.15	373	I	M	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y29513 +	7.00	93.19	210.84	705
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56774 +	7.00	97.54	225.88	373	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1997.DAT.W30749 -	7.00	93.12	210.61	712
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P90508 -	7.00	97.50	225.75	379	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y56971 +	7.00	92.87	209.75	739
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P90510 -	7.00	97.48	225.69	380	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.R97199 -	7.00	92.74	209.35	752
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P90511 -	7.00	97.48	225.69	380	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.R97199 -	7.00	91.96	206.76	842
/cgn2_2/cgdata/geneseq/geneseq/AA1989.DAT.P90512 -	7.00	97.48	225.69	380	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.Y39872 -	7.00	91.86	206.44	854
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80958 +	7.00	97.44	225.56	382	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R45779 +	7.00	91.38	204.85	916
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y73428 -	7.00	97.39	225.36	385	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y67882 -	7.00	91.20	204.29	939
/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.R97659 -	7.00	97.28	224.98	391	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R49039 +	7.00	91.04	203.77	961
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y50335 +	7.00	97.28	224.98	391	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y44717 -	7.00	90.97	203.54	971
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W51408 +	7.00	97.23	224.79	394	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W53280 -	7.00	90.52	202.09	1036
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y15193 +	7.00	97.23	224.79	394	I	C	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W53279 -	7.00	90.27	201.29	1074
/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R66920 +	7.00	97.21	224.73	395	I	M	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W80359 +	7.00	90.14	200.86	1095
/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.W01952 +	7.00	97.21	224.73	395	I	M	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y33496 +	7.00	89.61	199.18	1182
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W88282 +	7.00	97.21	224.73	395	I	F	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W72225 -	7.00	88.56	195.91	1374
/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.W69223 +	7.00	97.18	224.60	397	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W69753 -	7.00	88.56	195.91	1374
/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.W01955 +	7.00	97.18	224.60	397	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W72122 -	7.00	88.51	195.75	1384
/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R66921 +	7.00	97.16	224.54	398	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W72117 -	7.00	88.45	195.56	1396
/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.W01953 +	7.00	97.16	224.54	398	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W72117 -	7.00	88.45	195.56	1396
/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R66922 +	7.00	97.14	224.48	399	I	M	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R58809 +	7.00	88.21	194.82	1445
/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.W01954 +	7.00	97.14	224.48	399	I	M	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W40109 -	7.00	87.11	191.44	1694
/cgn2_2/cgdata/geneseq/geneseq/AA1992.DAT.R27809 -	7.00	96.84	223.39	417	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.R92120 +	7.00	87.01	191.13	1719
/cgn2_2/cgdata/geneseq/geneseq/AA1993.DAT.R37683 -	7.00	96.84	223.39	417	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y58575 -	7.00	86.57	189.80	1832
/cgn2_2/cgdata/geneseq/geneseq/AA1993.DAT.R37685 -	7.00	96.84	223.39	417	I	S	/cgn2_2/cgdata/geneseq/geneseq/AA1993.DAT.R33547 +	7.00	86.07	188.32	1967
/cgn2_2/cgdata/geneseq/geneseq/AA1993.DAT.R41345 -	7.00	96.84	223.39	417	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W63139 +	7.00	86.07	188.31	1968
/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R77401 -	7.00	96.84	223.39	417	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R72607 +	7.00	85.49	186.60	2138
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W57580 -	7.00	96.84	223.39	417	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R71003 +	7.00	85.41	186.36	2163
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56768 +	7.00	96.74	223.04	423	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y14589 -	7.00	85.16	185.62	2243
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W5778 +	7.00	96.70	222.92	425	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y14586 -	7.00	85.14	185.55	2250
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y04941 +	7.00	96.67	222.81	427	I	M	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y14588 -	7.00	85.11	185.45	2261
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W62778 +	7.00	96.61	222.58	431	I	F	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y14587 -	7.00	85.09	185.39	2268
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W62781 +	7.00	96.61	222.58	431	I	F	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W67738 -	7.00	84.48	183.33	2476
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56776 +	7.00	96.53	222.30	436	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1995.DAT.R71007 +	7.00	84.38	183.33	2510
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W98654 +	7.00	96.31	221.52	450	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W72204 -	7.00	82.87	179.00	3119
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y04948 +	7.00	96.31	221.52	450	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y08699 -	7.00	82.83	178.87	3170
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80957 +	7.00	96.29	221.47	451	I	E	/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y50976 +	7.00	82.58	178.15	3256
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56769 +	7.00	96.19	221.09	458	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y77179 +	7.00	81.54	175.24	3782
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56772 +	7.00	96.19	221.09	458	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1996.DAT.R97246 +	7.00	80.37	171.06	4472
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W56773 +	7.00	96.19	221.09	458	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R47861 -	7.00	80.26	170.74	4544
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.Y05442 +	7.00	96.19	221.09	458	I	H	/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R60517 -	7.00	80.26	170.74	4544
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y53635 -	7.00	96.07	220.67	466	I	A	/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W52846 -	7.00	79.51	168.60	5069
/cgn2_2/cgdata/geneseq/geneseq/AA1998.DAT.W63698 +	7.00	95.92	220.16	476	I	H					
/cgn2_2/cgdata/geneseq/geneseq/AA1997.DAT.W22849 +	7.00	95.86	219.95	480	I	C					
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80956 +	7.00	95.86	219.95	480	I	H					
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y80964 +	7.00	95.86	219.95	480	I	H					
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y15191 +	7.00	95.86	219.95	480	I	H					
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y10769 +	7.00	95.70	219.41	491	I	S					
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y70770 +	7.00	95.70	219.41	491	I	S					
/cgn2_2/cgdata/geneseq/geneseq/AA1993.DAT.R37682 -	7.00	95.48	218.63	507	I	S					
/cgn2_2/cgdata/geneseq/geneseq/AA2000.DAT.Y81602 +	7.00	95.48	218.63	507	I	S					
/cgn2_2/cgdata/geneseq/geneseq/AA1994.DAT.R58659 +	7.00	95.24	217.80	525	I	H					
/cgn2_2/cgdata/geneseq/geneseq/AA1999.DAT.W73473 +	7.00	95.19	217.61	529	I	H					

seq\_name: /cgn2\_2/cgdata/geneseq/geneseq/AA2000.DAT.Y83150

seq\_documentation\_block:

ID Y83150 standard; Protein: 465 AA.

XX

AC Y83150;

XX

DT 24-JUL-2000 (first entry)

XX

DE NGSP polypeptide of Neisseria gonorrhoeae.

XX NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
 KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
 KW screening; probe; primer; prophylaxis; therapy.  
 XX Neisseria gonorrhoeae.  
 OS  
 PN W0200012133-A1.  
 XX PD 09-MAR-2000.  
 XX 01-SEP-1999; 99WO-US20070.  
 XX 01-SEP-1998; 98US-0098685.  
 PR (ANTE-) ANTEX BIOLOGICS INC.  
 XX  
 PA Jackson WJ, Harris AM;  
 XX WPI; 2000-237782/20.  
 DR N-PSDB; 293414.  
 XX  
 PS Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
 PT Neisseria useful for diagnosis, prevention or treatment of Neisseria  
 PT infections  
 XX

PS Claim 5; Page 61-62; 68pp; English.

XX The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
 CC Arg-Gly-Asn motifs near the C-terminus which function as adherence  
 CC domains for extracellular matrix proteins. Using the NGSP polypeptide  
 CC as a vaccine produces antibodies which inhibit binding of N.  
 CC gonorrhoeae to the host's cellular matrix reducing attachment and/or  
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments  
 CC can be used to immunise an animal and produce an immune response.  
 CC They can also be used as ligands to detect antibodies elicited in  
 CC response to Neisseria infections and also as antigens or immunogens  
 CC for inducing Neisseria-specific antibodies which are useful in  
 CC immunoassays to detect Neisseria in biological specimens. Nucleotides  
 CC encoding NGSP or its fragments can be used as probes to identify  
 CC Neisseria in biological specimens by hybridization or polymerase  
 CC chain reaction amplification. The NGSP polypeptide can also be used  
 CC in screening assays to identify agents and compounds which useful as  
 CC diagnostic, prophylactic or therapeutic agents against Neisseria  
 CC infection.  
 XX

SQ Sequence 465 AA;

alignment\_scores:  
 Quality: 465.00 Length: 465  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y83150 ..

Align seg 1/1 to: Y83150 from: 1 to: 465

1 GTGTTCAAAAATACCAATACTTCGCTTTGGCGGCACTGTGTGCGCGCTT 50  
 |||||  
 1 ValPheLysLysTyrGlnTyrPheAlaLeuAlaLeuAlaLeuCysAlaAlaLe 17  
 |||||  
 51 GCTGCGAGCTCGGAAAAGCGAGCGAGCTTTTTCGCTGCGGAGCAAAAAAG 100  
 |||||  
 17 uLeuAlaGlyCysGluLysAlaGlySerPheGlyAlaAspLysLysG 34  
 |||||  
 101 AAGCATCTTCGTAGACGATCGAACACCAACAGACGAGCGCATGTC 150  
 |||||  
 34 luAlaSerPheValGluArgIleGluHisThrLysAspAspGlyServal 50  
 |||||  
 151 AGTATGCTGCTCGCGGCACTTTCGCAACTGGTTCAAGCGAAGGCGCGCG 200  
 |||||

51 SerMetLeuLeuProAspPheAlaGlnLeuValGlnSerGluGlyProAl 67  
 201 AGTCGTCAATATTTCAGCGAGCGCCCGCGCCGACCAACGAGCGAGCG 250  
 |||||  
 67 aValValAsnIleGlnAlaAlaProAlaProArgThrGlnAsnGlySerG 84  
 |||||  
 251 GCATGTCGGAACCGATTCCGACCGCTTGGCGACGAGCGACCGCTTCTAC 300  
 |||||  
 84 LyAsnAlaGluThrAspSerAspProLeuAlaAspSerAspProPheTyr 100  
 |||||  
 301 GAATTTTTCAAACGCTCGTCCGGAACATGCCGAAATCCGCCAAGAAGA 350  
 |||||  
 101 GluPhePheLysArgLeuValProasnMetProGluIleProGlnGlu 117  
 |||||  
 351 AGCAGATGACGCGGATTGAATTCGGTTCGGGCTTCATCATCAGCAAAA 400  
 |||||  
 117 uAlaAspAspGlyGlyLeuAsnPheGlySerGlyPheIleIleSerLysA 134  
 |||||  
 401 ACGGCTACATCCTGACCAATACCCACGCTGTTGCCGGTATGGGAGTATC 450  
 |||||  
 134 snGlyTyrIleLeuThrAsnThrHisValValAlaGlyMetGlySerIle 150  
 |||||  
 451 AAAGTCTGCTCAACGACGCGGAATATACCGCAAACTCATCGGTTC 500  
 |||||  
 151 LysValLeuLeuAsnAspLysArgGluTyrThrAlaLysLeuIleGlySe 167  
 |||||  
 501 GGATGTCCTCAATCCGATGTCGCCCTTCTGAAATCGACGCAACGAGAGC 550  
 |||||  
 167 rAspValGlnSerAspValAlaLeuLeuLysIleAspAlaThrGluLul 184  
 |||||  
 551 TACCCGTCGTCAAAATCGGCAATCCCAAAATTTGAAACCGGCGGAATGG 600  
 |||||  
 184 euProValValLysIleGlyAsnProLysAsnLeuLysProGlyGluTrp 200  
 |||||  
 601 GTCTCTGTCATCGCGCGCCCTTCGGCTTTGACACAGCGTCACCGCGCG 650  
 |||||  
 201 ValAlaAlaIleGlyAlaProPheGlyPheAspAsnSerValThrAlaGl 217  
 |||||  
 651 CATCTGTCTCCCAAGGAGGAGGAGGCTGCCCAACGAAAGCTACACACCT 700  
 |||||  
 217 yIleValSerAlaLysGlyArgSerLeuProAsnGluSerTyrThrProp 234  
 |||||  
 701 TCATCCAAACGAGTTCGCCATCAATCCGGGCAATTCGGCGCGCGCGCTG 750  
 |||||  
 234 heIleGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeu 250  
 |||||  
 751 TTCAACTTAAAGGACAGTCTCGGCATCAATTCGCAAAATATACAGCGG 800  
 |||||  
 251 PheAsnLeuLysGlyGlnValValGlyIleAsnSerGlnIleTyrSerAr 267  
 |||||  
 801 CAGCGCGGATTCATGGGCATCTCCTTTGCCATCCCGATTGACGTTGCCA 850  
 |||||  
 267 gSerGlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaLe 284  
 |||||  
 851 TGAATGTCGCGGACAGCTGAAACACCGGCAAAAGTCCCAACGCGGACAA 900  
 |||||  
 284 etAsnValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGln 300  
 |||||  
 901 CTGGCGGTGATTATTTCAGGAAGTACCTACGGTTTGGCACAGCTCGTTCG 950  
 |||||  
 301 LeuGlyValIleIleGlnGluValSerTyrGlyLeuAlaGlnSerPheG1 317  
 |||||  
 951 TCTGGATAAAGCCAGCGCGCATTTGATTGCCAAATCTCTCCCGGCGAGC 1000  
 |||||  
 317 yLeuAspLysAlaSerGlyAlaLeuIleAlaLysIleLeuProGlySerP 334  
 |||||  
 1001 CCGGAGAACGTCGCGGCTGCAGGCGGCGGACATCGTCTCAGCCTCGAC 1050  
 |||||  
 334 roAlaGluArgAlaGlyLeuGlnAlaGlyAspIleValLeuSerLeuAsp 350  
 |||||  
 1051 GGCGGAGAAATACGTTCTTCGGCGGCGACCTTCCTCGTCATGTCGGCGCAT 1100  
 |||||  
 351 GlyGlyGluIleArgSerSerGlyAspLeuProValMetValGlyAlaI1 367

```

1101 TAGCGCGGGAAGAAAGTACGCTCGCGGTATGGCGCAAGGCGAAGAA 1150
|||||
367 eThrProGlyLysGluValSerLeuGlyValTrpArgLysGlyGluGlu 384
1151 TCACATCAAGCCAGCTGGGAACGCCCGAGCATACCGGGCATCA 1200
|||||
384 leThrIleLysAlaLysLeuGlyAsnAlaAlaGluHisThrGlyAlaSer 400
1201 TCCAAACAGATCAAGCCCTCACCGAACAGCAATCCGGTACGTTCTC 1250
|||||
401 SerLysThrAspGluAlaProTyrThrGluGlnGlnSerGlyThrPheSe 417
1251 GGTCTGAATCCGAGGATACCTTCCAGACACATACCGACAGCGGCA 1300
|||||
417 rValGluSerAlaGlyIleThrLeuGlnThrHisThrAspSerSerGlyL 434
1301 AACACCTCGTCGTCAGCGGTTTCCGACGCGGAGACGCGAGGCTTA 1350
|||||
434 yHisLeuValValValArgValSerAspAlaAlaGluArgAlaGlyLeu 450
1351 AGCAGCGGCGACAAATCCCTAGCGTCAGGCGCAAGTCCCGTCAA 1395
|||||
451 ArgHisGlyAspGluLeuAlaValArgAlaSerProArgGln 465

```

seq\_name: /cgn\_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75748

seq\_documentation\_block:  
ID\_Y75748 standard; Protein; 499 AA.

```

XX AC Y75748;
XX DT
XX DE Neisseria gonorrhoe ORF 986 protein sequence SEQ ID NO:2968.
XX KW
XX KW Neisseria meningitidis; Neisseria gonorrhoe; antigen; vaccine;
XX KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
XX KW antibacterial; gene therapy.
XX OS
XX OS Neisseria gonorrhoe.
XX PN
XX PN W09957280-A2.
XX PD
XX PD 11-NOV-1999.
XX PF
XX PF 30-APR-1999; 99WO-US09346.
XX PR
XX PR 01-MAY-1998; 98US-0083758.
XX PR 31-JUL-1998; 98US-0094869.
XX PR 02-SEP-1998; 98US-0098994.
XX PR 02-SEP-1998; 98US-0099062.
XX PR 09-OCT-1998; 98US-0103749.
XX PR 09-OCT-1998; 98US-0103794.
XX PR 09-OCT-1998; 98US-0103796.
XX PR 25-FEB-1999; 99US-0121528.
XX PA
XX PA (CHIR ) CHIRON CORP.
XX PA (GENO-) INST GENOMIC RES.
XX PI
XX PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
XX PI Petersen J, Pizsa M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
XX PI Tettelin H, Venter JC;
XX XX
XX DR WPI; 2000-062150/05.
XX DR N-PSDB; 254510.
XX XX

```

Novel Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics -  
Claim 2; Page 1390; 1453pp; English.  
253015 to 254536, 254577 to 254615, and Y74253 to Y75941 represent

CC novel Neisseria meningitis and N. gonorrhoe polynucleotides and polypeptides. 254537 to 254576 and 254616 to 255473 represent PCR primers used in the exemplification of the present invention. The CC polypeptides, the polynucleotides, antibodies and compositions of CC the invention can be used as vaccines, as diagnostic reagents, and as CC immunogenic compositions. The polypeptides can be used in the CC manufacture of medicaments for treating or preventing infection due to CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the CC presence of Neisseria bacteria, or to raise antibodies. They may also CC be used to screen for agonists or antagonists, which may themselves CC have use as antibacterial agents. The polynucleotides of the invention CC may also be used in gene therapy protocols.

XX  
SQ Sequence 499 AA:

alignment\_scores:  
Quality: 451.00 Length: 451  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x Y75748

Align seg 1/1 to: Y75748 from: 1 to: 499

```

1 GTGTTCAAAAAATACCAATACTTTCGGCGCACACTGTGCGCGCTT 50
|||||
1 ValPheLysLysTyrGlnTyrPheAlaLeuAlaAlaLeuCysAlaAlaLe 17
51 GCTGCAGGCTCGGAAAGGAGCGAGCTTTTTCGGTGGCGGACAAAG 100
|||||
17 uLeuAlaGlyCysGluLysAlaGlySerPhePheGlyAlaAspLysLysG 34
101 AAGCATCTTCGTAGACGCATCGAACACACCAAGACGACGCGAGTGC 150
|||||
34 uAlaSerPheValGluArgIleGluHisThrLysAspAspGlySerVal 50
151 AGTATGCTGCTCCCGGACCTTTCGCCCAACTGGTTCAAAGGGAAGCGCCGC 200
|||||
51 SerMetLeuLeuProAspPheAlaGlnLeuValGlnSerGluGlyProAl 67
201 AGTCGTCAATATTCAGGAGCGCCCGCGCGCACCCCAAAACGGGAGCG 250
|||||
67 aValValAsnIleGlnAlaAlaProAlaProArgThrGlnAsnGlySerG 84
251 GCAATGCCGAAACCGATTCCGACCCCGCTTTCGCCACAGCGACCGCTTCTAC 300
|||||
84 LysAsnAlaGluThrAspSerAspProLeuAlaAspSerAspProPheTyr 100
301 GAATTTTTCAAACGGCTCGTCCCGAACATCCCGGAAATCCCAAGAACGA 350
|||||
101 GluPhePheLysArgLeuValProAsnMetProGluIleProGlnGluG 117
351 AGCAGATGACGGCGGATTCGAATTCGGTTCGGGCTTCATCATCAGCAAAA 400
|||||
117 uAlaAspAspGlyGlyLeuAsnPhgGlySerGlyPheIleIleSerLysA 134
401 AGCGCTACATCTCGACCAATACCGCTGTTGCGCGGTATGGCGAGTATC 450
|||||
134 snGlyTyrIleLeuThrAsnThrHisValValAlaGlyMetGlySerIle 150
451 AAAGTCCTGCTCAACGACACAGCGCGCAATATACCGCCCAAACTCATCGTTC 500
|||||
151 LysValLeuLeuAsnAspLysArgGluTyrThrAlaLysLeuIleGlySe 167
501 GGATGTCNAATCCGATGTCGCCCTTCCTGAAATCGACGCAACGGAAGAGC 550
|||||
167 rAspValGlnSerAspValAlaLeuLeuLysIleAspAlaThrGluGluL 184
551 TACCCGTCGTCAAAATCCGCAATCCCAAAATTTGAAACCGCGCGGAATGG 600
|||||
184 euProValValLysIleGlyAsnProLysAsnLeuLysProGlyGluTrp 200

```



```
601 GTCGCTGCCATCGCGCGCCCTTCGGCTTTGACAAACAGCGTGACCGCCGG 650
|||||
201 ValAlaIleIleGlyAlaProPheGlyPheAspAsnSerValThrAlaG1 217
|||||
651 CATCGTGTCCGCCAAAGCAGAGCGCTGCCCAACGAAAGCTACACACCGT 700
|||||
217 yIleValSerAlaLysGlyArgSerLeuProAsnGluSerTyrThrProp 234
|||||
701 TCATCCAAACGACGCTGCCATCAATCCGGGCAATTCGGCGCCCGCTG 750
|||||
234 heIleGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeu 250
|||||
751 TTCAACTTAAAGGACAGCTGCTCGGCATCAATTCGCAAAATATACAGCG 800
|||||
251 PheAsnLeuLysGlyGlnValValGlyIleAsnSerGlnIleTyrSerAr 267
|||||
801 CAGCGCGGATTCATGGGCATCTCCTTTGGCCATCCGATTCGACGTTGCCA 850
|||||
267 gSerGlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaM 284
|||||
851 TGAATGTCCGCGAAGCAGCTGAAACACCGGCAAGTCCCAACGCGGACAA 900
|||||
284 etAsnValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGln 300
|||||
901 CTGGCGGTGATTATTCAGAAGTATCCTACGGTTTGGGCACAGTCGTTCCG 950
|||||
301 LeuGlyValIleIleGlnGluValSerTyrGlyLeuAlaGlnSerPheG1 317
|||||
951 TCTGGATAAGCCAGCGCGCATTCATTCGCAAAATCCTCCGCGCAGCC 1000
|||||
317 yLeuAspLysAlaSerGlyAlaLeuIleAlaLysIleLeuProGlySerP 334
|||||
1001 CCGCAGAACGTCGCCGCTGCGAGCGGCGACATCGTCTCAGCGCTCGAC 1050
|||||
334 roAlaGluArgAlaGlyLeuGlnAlaGlyAspIleValLeuSerLeuAsp 350
|||||
1051 GCGGAGAAATACGTTCTTCGCGCGACCTTCCGTCATCGTGGCGCCAT 1100
|||||
351 GlyGlyGluIleArgSerSerGlyAspLeuProValMetValGlyAlaI1 367
|||||
1101 TACCGCGGAAAGAAAGTCAGCTCGGCGTATGGCGCAAGCGGAAGAAA 1150
|||||
367 eThrProGlyLysGluValSerLeuGlyValIlePArgLysGlyGluGlu 384
|||||
1151 TCACAATCAAGCCAAAGCTGGGCAACGCGCGAGCATACCGCGCGCATCA 1200
|||||
384 leThrIleLysAlaLysLeuGlyAsnAlaAlaGluHisThrGlyAlaSer 400
|||||
1201 TCCAAACAGATGAAGCCCTTACCGCAACAGCAATCCGGTACGTTCTC 1250
|||||
401 SerLysThrAspGluAlaProTyrThrGluGlnGlnSerGlyThrPheSe 417
|||||
1251 GGTGCAATCCGAGCATTACCTTTCAGACACATACCGACAGCAGCGCA 1300
|||||
417 rValGluSerAlaGlyIleThrLeuGlnIleThrHisThrAspSerSerGlyL 434
|||||
1301 AACACCTCGTGTACGGTTTCCGAGCGGCGAGAACGCGCAGGCTTA 1350
|||||
434 yHisLeuValValValArgValSerAspAlaAlaGluArgAlaGlyLeu 450
|||||
1351 AGG 1353
|||
451 Arg 451
```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT.Y83151  
seq\_documentation\_block:  
ID Y83151 standard; Protein: 414 AA.  
XX  
AC Y83151;  
XX  
DT 24-JUL-2000 (first entry)

```
XX NGSP polypeptide of Neisseria gonorrhoeae.  
DE  
XX NGSP; polypeptide; peptide; vaccine; immune response; antibody;  
KW cellular matrix; adherence domain; ligand; detection; diagnosis;  
KW screening; probe; primer; prophylaxis; therapy.  
XX Neisseria gonorrhoeae.  
OS  
XX  
XX WO200012133-A1.  
PN  
XX 09-MAR-2000.  
PD  
XX 01-SEP-1999; 99WO-US20070.  
PF  
XX 01-SEP-1998; 98US-0098685.  
PR  
XX (ANTE-) ANTEX BIOLOGICS INC.  
PA  
XX Jackson WJ, Harris AM;  
PI  
XX  
XX WPI; 2000-237782/20.  
DR  
XX N-PSDB; Z93415.  
DR  
XX Non-cytosolic NGSP polypeptide and polynucleotide sequence from  
PT Neisseria useful for diagnosis, prevention or treatment of Neisseria  
PT infections  
PT  
XX  
PS Claim 5; Page 63-64; 68pp; English.  
XX  
XX The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and  
CC Arg-Gly-Asn motifs near the C-terminus which function as adherence  
CC domains for extracellular matrix proteins. Using the NGSP polypeptide  
CC as a vaccine produces antibodies which inhibit binding of N.  
CC gonorrhoeae to the host's cellular matrix reducing attachment and/or  
CC subsequent invasion. The NGSP polypeptide and its peptide fragments  
CC can be used to immunise an animal and produce an immune response.  
CC They can also be used as ligands to detect antibodies elicited in  
CC response to Neisseria infections and also as antigens or immunogens  
CC for inducing Neisseria-specific antibodies which are useful in  
CC immunoassays to detect Neisseria in biological specimens. Nucleotides  
CC encoding NGSP or its fragments can be used as probes to identify  
CC Neisseria in biological specimens by hybridization or polymerase  
CC chain reaction amplification. The NGSP polypeptide can also be used  
CC in screening assays to identify agents and compounds which useful as  
CC diagnostic, prophylactic or therapeutic agents against Neisseria  
CC infection.  
XX  
XX Sequence 414 AA;
```

alignment\_scores:  
Quality: 414.00 Length: 414  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000  
..  
alignment\_block:  
US-09-388-090-3 x Y83151 ..  
Align seg 1/1 to: Y83151 from: 1 to: 414  
154 ATGCTGCTGCCGCGACTTTGCCAACTGGTTCAAAGCGAAGCGCCGCGAGT 203  
|||||  
1 MetLeuLeuProAspPheAlaGlnLeuValGlnSerGluGlyProAlaVal 17  
204 CGTCAATATTTCAGGCAGCGCCCGCGCGCGCACCCCAACGCGCAGCGCA 253  
|||||  
17 lValAsnIleGlnAlaAlaProAlaProArgThrGlnAsnGlySerGlyA 34  
254 ATGCCGAAACCGATTCCGACCGCGCTTCCGACAGCGACCGCTTCTACGAA 303  
|||||  
34 snAlaGluThrAspSerAspProLeuAlaAspSerAspProPheTyrGlu 50

```
304 TTTTTCACACGGCCTCGTCCGGAACATGCCGAATCCCCCAAGAAGACG 353
51 PhePheLysArgLeuValProAsnMetProGluIleProGlnGluGluAl 67
354 AGATGACGGCGGATGAACCTCGGTTCGGGCTTCATCATCAGCAAAACG 403
67 aspaspGlyLeuAsnPheGlySerGlyPheIleIleSerLysAsnG 84
404 GCTACATCTGTACCAATACCCAGCTGTTCCCGGTATGGCAGTATCAAA 453
84 LyTyrIleLeuThrAsnThrHisValValAlaGlyMetGlySerIleLys 100
454 GTCCTGCTCAACGACAGCGCGAATATACCGCAAACTCATCGGTCCGA 503
101 ValLeuLeuAsnAspLysArgGluTyrThrAlaLysLeuIleGlySerAs 117
504 TGTCCCAATCCGATGTCGCGCTTCTGAAATCGACGCAACGAGAGCTAC 553
117 pValGlnSerAspValAlaLeuLeuLysIleAspAlaThrGluGluLeuP 134
554 CGTCTGCAAAATCGGCAATCCCAAAATTTGAAACCGGCGAATGGGTC 603
134 roValValLysIleGlyAsnProLysAsnLeuLysProGlyGluTrpVal 150
604 GCTGCCATCGCGCGCCCTTCGGCTTTGACAACAGCGTGACCGCGGCAT 653
151 AlaAlaIleGlyAlaProPheGlyPheAspAsnSerValThrAlaGlyI 167
654 CGTGTCCGCAAGGAGGAGCGCTGCCCAACGAAAGCTACACACCCCTCA 703
167 eValSerAlaLysGlyArgSerLeuProAsnGluSerTyrThrProPheI 184
704 TCCAAACCGACGTTGCCATCAATCCGGGCAATTCGGCGCGCCGCTGTC 753
184 IeGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeuPhe 200
754 AACTTAAAGGACAGCTGCTGGCGCATCAATTCGCAAAATATACAGCCG 803
201 AsnLeuLysGlyGlnValValGlyIleAsnSerGlnIleTyrSerArgSe 217
804 CGCGGATTATCGGCGATCTCCTTTGCCATCCCGATTGACGTTGCCATGA 853
217 rGlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaMetA 234
854 ATGTCCCGCAACAGCTGAAACACCGCGGCAAGTCCAAACGCGGACAACTG 903
234 snValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGlnLeu 250
904 GCGCTGATTATTCAGGAAGTATCTACGGTTTGGCAGAGTCGTTCCGTCT 953
251 GlyValIleIleGlnGluValSerTyrGlyLeuAlaGlnSerPheGlyLe 267
954 GGATAAGCCAGCGCGCATTTGATTCGCAAAATCCTTCGCGCGACGCCG 1003
267 uAspLysAlaSerGlyAlaLeuIleAlaLysIleLeuProGlySerProA 284
1004 CAGAAGTCGCGGCTTCGAGCGGGGACATCGTCTCTCAGCCCTCGACGGC 1053
284 leGluArgAlaGlyLeuGlnAlaGlyAspIleValLeuSerLeuAspGly 300
1054 GGAGAATACGTTCTTCGGCGACCTTCGCGTCATGTCGGCGGCATTAC 1103
301 GlyGluIleArgSerSerGlyAspLeuProValMetValGlyAlaIleTh 317
1104 GCGGGAAAGAGTCAGCTCGCGCTATGGCGCAAGGCGAAGAAATCA 1153
317 rProGlyLysGluValSerLeuGlyValTrpArgLysGlyGluGluIle 334
1154 CAATTAAGCCAGCTGGCAACCGCGCGGATACCGCGGCATCATCC 1203
334 hrIleLysAlaLysLeuGlyAsnAlaAlaGluHisThrGlyAlaSerSer 350
1204 AAAACAGATGAAGCCCTTACACCGAACCAATCCGGTACGTTCTCGGT 1253
```

```
|||||
351 LysThrAspGluAlaProTyrThrGluGlnGlnSerGlyThrPheSerVa 367
1254 CGAATCCGCGAGGCATTACCTCTCAGACACATACCGACAGCGGCAAAAC 1303
367 IGlSerAlaGlyIleThrLeuGlnThrHisThrAspSerSerGlyLysH 384
1304 ACCTCGTCTGTCGTACGGGTTTCCGCGCGCAGACGCGCAGGCTTAAAGG 1353
384 IsLeuValValValArgValSerAspAlaAlaGluArgAlaGlyLeuArg 400
1354 CACGCGGAGGAATCTAGCCGTCAGGGCAAGTCCCGCTCAA 1395
401 HisGlyAspGluIleLeuAlaValArgAlaSerProArgGln 414

seq_name: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52996

seq_documentation_block:
ID Y52996 standard; Protein; 370 AA.
XX
AC Y52996;
XX
DT 21-FEB-2000 (first entry)
XX
DE Neisseria meningitidis BASB013-C protein sequence.
XX
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
KW invasive bacterial disease; antibacterial.
XX
OS Neisseria meningitidis.
XX
PN WO9955872-A1.
XX
PD 04-NOV-1999.
XX
PF 20-APR-1999; 99WO-EP02765.
XX
PR 23-APR-1998; 98GB-0008734.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Ruelle J;
XX
DR WPI: 2000-052809/04.
DR N-PSDB; Z33308.
XX
PT Novel polynucleotides and polypeptides from Neisseria meningitis used
PS to prepare vaccines against bacterial infections -
XX
PS Example 2; Page 82-83; 94pp; English.
XX
CC The present sequence is a conserved BASB013-C polypeptide isolated
CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides
CC may be employed as research reagents and material for the discovery of
CC treatments and diagnostics for diseases, particularly human diseases.
CC They can be used for diagnosis of disease, staging of disease, or
CC determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and
CC for screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB013 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies. The
CC polypeptides can also be used in vaccine formulations, and to identify
CC agonists and antagonists. The polypeptides, antibodies, agonists and
CC antagonists (which are bacteriostatic) are used for the treatment and
CC prevention of diseases such as upper respiratory tract infection,
CC invasive bacterial diseases such as bacteraemia and meningitis, and for
CC in the prevention and screening of bacteria to eukaryotic matrix proteins
CC on in-dwelling devices, or to extracellular proteins on wounds, and to
CC thus prevent tissue damage and/or block the normal progression of
CC pathogenesis in infections initiated other than by the implantation of
CC
```

CC in-dwelling devices or by other surgical techniques.

XX Sequence 370 AA;

alignment\_scores:  
Quality: 127.00 Length: 127  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y52996 ..

Align seg 1/1 to: Y52996 from: 1 to: 370

```
583 TTGAACCGCGGAATGGTCCTGCCATCGCGCGCCCTTCGGCTTTGA 632
|||||
195 LeuLysProGlyGluTrpValAlaAlaIleGlyAlaProPheGlyPheAs 211
|||||
633 CAACAGCGTGACCGCGCGCATCGTCTCCGCCAAAGGCAGACGCTGCCA 682
|||||
211 pAsnSerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProA 228
|||||
683 ACAGAAAGCTACACACCCCTTCATCCAAACGACGTTGCCATCAATCGGGC 732
|||||
228 snGluSerTyThrProPheIleGlnThrAspValAlaIleAsnProGly 244
|||||
733 AATTCCGCGCGCGCTCTTCAACTTAAAGGACAGGTCGTCGGCATCAA 782
|||||
245 AsnSerGlyGlyProLeuPheAsnLeuLysGlyGlnValValGlyIleAs 261
|||||
783 TTCGCAATATACACCGCGCGGGATTATCGGCGCATCTCTTTGCCA 832
|||||
261 nSerGlnIleTySerArgSerGlyGlyPheMetGlyIleSerPheAlaI 278
|||||
833 TCCGATTGACGTTCCCATGAATGTCGCGAAGACGCTGAAACACCGGC 882
|||||
278 leProIleAspValAlaMetAsnValAlaGluGlnLeuLysAsnThrGly 294
|||||
883 AAAGTCCACCGCGGACAACTCGGCGTGATTATTCAGGAAGTATCCTTACGG 932
|||||
295 LysValGlnArgGlyGlnLeuGlyValIleIleGlnGluValSerTyrgL 311
|||||
933 TTTGGCACAGTCGTTCGTCGTGGGATAAAGCC 963
|||||
311 yLeuAlaGlnSerPheGlyLeuAspLysAla 321
|||||
```

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqp/AA2000.DAT:Y5749

seq\_documentation\_block:

ID Y5749 standard; Protein; 499 AA.

XX Y5749;

XX

XX

DT 21-MAR-2000 (first entry)

XX

DE Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2970.

XX

KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;

XX

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

XX

XX antibacterial; gene therapy.

OS Neisseria meningitidis.

XX

PN W09957280-A2.

XX

PD 11-NOV-1999.

XX

PF 30-APR-1999; 99WO-US09346.

XX

PR 01-MAY-1998; 98US-0083758.

XX

PR 31-JUL-1998; 98US-0094869.

XX

PR 02-SEP-1998; 98US-0098994.

PR 02-SEP-1998; 98US-0099062.  
PR 09-OCT-1998; 98US-0103749.  
PR 09-OCT-1998; 98US-0103794.  
PR 09-OCT-1998; 98US-0103796.  
PR 25-FEB-1999; 99US-0121528.  
XX (CHIR ) CHIRON CORP.  
PA (GENO-) INST GENOMIC RES.  
XX  
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;  
PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;  
PI Tettelin H, Venter JC;  
XX  
DR WPI; 2000-062150/05.  
DR N-PSDB; Z54511.  
XX  
PT Novel Neisserial polypeptides predicted to be useful antigens for  
PT vaccines and diagnostics  
XX  
PS Claim 2: Page 1390; 1453pp; English.  
XX  
CC Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent  
CC novel Neisseria meningitis and N. gonorrhoeae polynucleotides and  
CC polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR  
CC primers used in the exemplification of the present invention. The  
CC polypeptides, the polynucleotides, antibodies and compositions of  
CC the invention can be used as vaccines, as diagnostic reagents, and as  
CC immunogenic compositions. The polypeptides can be used in the  
CC manufacture of medicaments for treating or preventing infection due to  
CC Neisserial bacteria (e.g. meningitis and septicaemia), to detect the  
CC presence of Neisseria bacteria, or to raise antibodies. They may also  
CC be used to screen for agonists or antagonists, which may themselves  
CC have use as antibacterial agents. The polynucleotides of the invention  
CC may also be used in gene therapy protocols.  
XX  
SQ Sequence 499 AA;

alignment\_scores:

Quality: 127.00 Length: 127  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y5749 ..

Align seg 1/1 to: Y5749 from: 1 to: 499

```
583 TTGAACCGCGGAATGGTCGTCGCCATCGCGCGCCCTTCGGCTTTGA 632
|||||
195 LeuLysProGlyGluTrpValAlaAlaIleGlyAlaProPheGlyPheAs 211
|||||
633 CAACAGCGTGACCGCGCGCATCGTCTCCGCCAAAGGCAGACGCTGCCA 682
|||||
211 pAsnSerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProA 228
|||||
683 ACAGAAAGCTACACACCCCTTCATCCAAACGACGTTGCCATCAATCGGGC 732
|||||
228 snGluSerTyThrProPheIleGlnThrAspValAlaIleAsnProGly 244
|||||
733 AATTCCGCGCGCGCTCTTCAACTTAAAGGACAGGTCGTCGGCATCAA 782
|||||
245 AsnSerGlyGlyProLeuPheAsnLeuLysGlyGlnValValGlyIleAs 261
|||||
783 TTCGCAATATACACCGCGCGCATCGTCTCCGCCAAAGGCAGACGCTGCCA 832
|||||
261 nSerGlnIleTySerArgSerGlyGlyPheMetGlyIleSerPheAlaI 278
|||||
833 TCCCGATTGACGTTCCCATGAATGTCGCGAAGACGCTGAAACACCGGC 882
|||||
278 leProIleAspValAlaMetAsnValAlaGluGlnLeuLysAsnThrGly 294
|||||
883 AAAGTCCACCGCGGACAACTCGGCGTGATTATTCAGGAAGTATCCTTACGG 932
```

|||||  
295 LysValGlnArgGlyGlnLeuGlyVallelleGlnGluValSerTyrG1 311  
933 TTTCGCACAGTCGTCGCTCGGATAAGCC 963  
|||||  
311 yLeuAlaGlnSerPheGlyLeuAspLysAla 321

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52994

seq\_documentation\_block:  
ID Y52994 standard; Protein; 499 AA.

AC Y52994;  
DT 21-FEB-2000 (first entry)

DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.  
XX  
XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial.  
XX  
OS Neisseria meningitidis.

PN WO9955872-AL.  
XX  
XX 04-NOV-1999.  
XX  
XX 20-APR-1999; 99WO-EP02765.  
XX  
XX 23-APR-1998; 98GB-0008734.  
PR  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
XX Ruelle J;

XX  
XX WPI; 2000-052809/04.  
DR  
XX N-PSDB; 233306.  
XX  
PT Novel polynucleotides and polypeptides from Neisseria meningitidis used  
PT to prepare vaccines against bacterial infections  
XX  
XX Claim 3; Page 77-78; 94pp; English.

XX  
CC The present sequence represents a BASB013 polypeptide isolated from  
CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be  
CC employed as research reagents and material for the discovery of  
CC treatments and diagnostics for diseases, particularly human diseases.  
CC They can be used for diagnosis of disease, staging of disease, or  
CC determining response of an infectious organism to drugs. The  
CC polynucleotides may be used as a source for hybridisation probes, and  
CC for screening of genetic mutations, serotype, organism or strain  
CC identification, identification of mutation in BASB013 sequences, and as  
CC components of arrays which are useful for diagnostic and prognostic  
CC purposes. The polypeptides can be used to produce antibodies. The  
CC polypeptides can also be used in vaccine formulations, and to identify  
CC agonists and antagonists. The polypeptides, antibodies, agonists and  
CC antagonists (which are bacteriostatic) are used for the treatment and  
CC prevention of diseases such as upper respiratory tract infection,  
CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
CC the development and screening of antibacterial drugs. They are also used  
CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
CC thus prevent tissue damage and/or block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or by other surgical techniques.

XX Sequence 499 AA;

alignment\_scores:  
Quality: 127.00 Length: 127  
Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x Y52994

Align seg 1/1 to: Y52994 from: 1 to: 499

583 TTGAAACGGCGCGAATGGTCTGCTGCCATCGCGCGCCCTTCGCGTTTGA 632  
|||||  
195 LeuLysProGlyGlyTrpValAlaAlaIleGlyAlaProPheGlyPheAs 211  
633 CAACAGCGTGACCGCGCATCGTCTGCCCAAAGGCAGAAAGCCCTGCCCA 682  
|||||  
211 pAsnSerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProA 228  
683 ACGAAAGCTACACACCCCTTCATCCAAACCGCAGCTTCCCATCAATCCGGGC 732  
|||||  
228 snGluSerTyrThrProPheIleGlnThrAspValAlaIleAsnProGly 244  
733 AATTCCGGCGCGCGCTCTTCAACTTAAAGGACAGGTCTCGGCATCAA 782  
|||||  
245 AsnSerGlyGlyProLeuPheAsnLeuLysGlyGlnValValGlyIleAs 261  
783 TTCCCAAATATACAGCCGCGCGGATTCATGGGCATCTCCTTTGCCA 832  
|||||  
261 nSerGlnIleTyrSerArgSerGlyGlyPheMetGlyIleSerPheAlaI 278  
833 TCCCGATGAGCTTCCCATGAATGTCGCGCAACAGCTGAAACACACCGGC 882  
|||||  
278 leProIleAspValAlaMetAsnValAlaGluGlnLeuLysAsnThrGly 294  
883 AAGTCCAACGCGGACAACTGGCGTGATTATTTCAGGAAGTATCCTACGG 932  
|||||  
295 LysValGlnArgGlyGlnLeuGlyValIleIleGlnGluValSerTyrG 311  
933 TTTCGCACAGTCGTCGCTCGGATAAGCC 963  
|||||  
311 yLeuAlaGlnSerPheGlyLeuAspLysAla 321

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52995

seq\_documentation\_block:  
ID Y52995 standard; Protein; 499 AA.

XX Y52995;  
XX  
XX 21-FEB-2000 (first entry)

XX  
XX DE Neisseria meningitidis strain H44/76 BASB013 protein sequence.  
XX  
XX KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;  
KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;  
KW invasive bacterial disease; antibacterial.

XX Neisseria meningitidis.

XX WO9955872-AL.

XX 04-NOV-1999.

XX PF 20-APR-1999; 99WO-EP02765.

XX PR 23-APR-1998; 98GB-0008734.

XX PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX Ruelle J;

XX WPI; 2000-052809/04.

XX N-PSDB; 233307.

XX Novel polynucleotides and polypeptides from Neisseria meningitidis used  
PT to prepare vaccines against bacterial infections



```

207 ProPheGlyPheAsnSerValThrAlaGlyIleValSerAlaLysGln 223
669 CAGAAAGCTGCCCAAGAAAGCTACACACCCCTTCATCCAAACCGACGTTG 718
|||||
223 yargSerLeuProAsnGluSerTyrThrProPheIleGlnThrAspValA 240
|||||
719 CCATCAATCCGGGCAATTCCGGGGGCCCGTGTCACTTAAAGGACAG 768
|||||
240 laIleAsnProGlyAsnSerGlyGlyProLeuPheAsnLeuLysGlyGln 256
|||||
769 GTCGTCGGCATCAATTCGCAAAATATACAGCCGCGACGCGGATTCATGGG 818
|||||
257 ValValGlyIleAsnSerGlnIleTyrSerArgSerGlyGlyPheMetGln 273
|||||
819 CATCTCCTTTGCCATCCGATGTGACGTTGCCATCAATGTCGCCGAACAGC 868
|||||
273 yIleSerPheAlaIleProIleAspValAlaMetAsnValIlaGluGlnL 290
|||||
869 TGAATAACACCGGCAAGTCCACGCGGACCAACTGGGCGTGTATTATTCAG 918
|||||
290 euLysAsnThrGlyLysValGlnArgGlyGlnLeuGlyValIleIleGln 306
|||||
919 GAAGTATCTCTACGTTTGGGCACAGCTCCGTCCTGGATAAGCC 963
|||||
307 GluValSerTyrGlyLeuAlaGlnSerPheGlyLeuAspLysAla 321
|||||
seq_name: /cqn2_2/cqndata/geneseq/geneseq/AA2000..DAT.Y70409

```

CC may be used for diagnosis, therapy or prophylaxis of Neisserial  
CC infections such as, bacterial meningitidis and septicaemia.  
CC Note: The protein represented in SEQ ID NO:2 of the specification is  
CC erroneous. The present sequence is the decoded version of the nucleotide  
CC represented in Z51533.

XX  
SQ Sequence 448 AA:

alignment\_scores:  
Quality: 103.00 Length: 103  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x Y70409 ..

Align seg 1/1 to: Y70409 from: 1 to: 448

655 GTCTCGCCCAAGGACAGAAGCTGCCAACGAAGCTACACACCCTTCA<sup>1</sup> 704  
|||||  
168 ValSerAlalysGlyArgSerLeuProAsnGluSerTyrThrProPhe<sup>1</sup> 184  
|||||  
705 CCAAACGCAGCTTGCCATCAATCCGGGCAATTCCGGCGCCCGCTGTTC<sup>1</sup> 754  
|||||  
184 eGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeuPhe<sup>1</sup> 201  
|||||  
755 ACTTAANGACAGGTGCTCGGCATCAATTCGCAATATACAGCCGCAGC<sup>1</sup> 804  
|||||  
201 snLeuLysGlyGlnValValcGlyIleAsnSerGlnIleTyrSerArgSer<sup>1</sup> 217  
|||||  
805 GGCGGATTCATGGCATCTCCTTTGCCATCCGATTTGACGTTGCCATGAA<sup>1</sup> 854  
|||||  
218 GlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaMetas<sup>1</sup> 234  
|||||  
855 TGTCGCCGAACAGCTGAAAAACACCGCAAGTCCAACCGGACAACTGG<sup>1</sup> 904  
|||||  
234 nValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGlnLeuG<sup>1</sup> 251  
|||||  
905 GCGTGATTATTCAGGAGTAGTATCCTACGGTTTGGCACAGTCGTTCCGTCTG<sup>1</sup> 954  
|||||  
251 llyValIleIleGlnGluValserTyrGlyLeuAlaGlnSerPheGlyLeu<sup>1</sup> 267  
|||||  
955 GATAAAGCC 963  
|||||  
268 AspLysAla 270

seq\_name: /cgndata/geneseq/geneseqp/AA2000.DAT:Y70414

seq\_documentation\_block:  
ID\_Y70414 standard; Protein: 475 AA.

XX  
AC Y70414;  
XX  
DT DT (first entry)  
XX  
DE Neisseria meningitidis NMSP protein-3.  
XX  
KW NMSP; non-cytosolic; antibacterial; anti-inflammatory; cytotoxic;  
KW anti-NMSP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia.  
XX  
OS Neisseria meningitidis.  
XX  
FH Key Location/Qualifiers  
FT Misc-difference 19 /note= "Encoded by GAA"  
FT FT Misc-difference 35 /note= "Encoded by GCC"  
FT FT Misc-difference 61 /note= "Encoded by GGC"  
FT FT Misc-difference 111 /note= "Encoded by AAC"

FT Misc-difference 122 /note= "Encoded by GCC"  
FT Misc-difference 171 /note= "Encoded by AAT"  
FT Misc-difference 194 /note= "Encoded by GGCATC"  
FT Misc-difference 298 /note= "Encoded by AGC"  
FT Misc-difference 317 /note= "Encoded by AGG"  
FT Misc-difference 364 /note= "Encoded by GCC"  
FT Misc-difference 373 /note= "Encoded by ACC"  
FT Misc-difference 410 /note= "Encoded by AAA"  
FT Misc-difference 428 /note= "Encoded by CAC"  
FT Misc-difference 436 /note= "Encoded by AGG"  
FT Misc-difference 437 /note= "Encoded by GCA"  
FT Misc-difference 438 /note= "Encoded by AGT"  
FT Misc-difference 440..475 /note= "Encoded by CGTCAA"

W0200012535-A2.

09-MAR-2000.

01-SEP-1999; 99WO-US19663.

01-SEP-1998; 98US-0098685.

(ANTE-) ANTEX BIOLOGICS INC.

Jackson WJ, Harris AM;

WPI; 2000-256581/22.

N-PSDB; 251539.

Neisseria meningitidis NMASP polypeptide, nucleotide sequences and antibodies, useful in vaccines against infection -

Claim 5; Page 71-73; 75pp; English.

The present sequence is the Neisseria meningitidis NMASP protein. NMASP is a non-cytosolic protein, with antibacterial and anti-inflammatory activity. It shows sequence similarity to E. coli DegP (HtrA) protein. NMASP proteins can be used as ligands to detect antibodies elicited in response to N. meningitidis infections. Cytotoxic anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA may be used for diagnosis, therapy or prophylaxis of Neisserial infections such as, bacterial meningitis and septicaemia.

Sequence 475 AA;

alignment\_scores:  
Quality: 103.00 Length: 103  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y70414 ..

Align seg 1/1 to: Y70414 from: 1 to: 475

655 GTGTCGCGCAAGGAGGAGCTGCCAACGAGTACACACCTTCAT 704  
|||||  
195 ValSerAlaLysGlyArgSerLeuProAsnGluSerTyrThrProPhe11 211

705 CCAAAACGACGTTGCCATCAATCCGGGCAATTCCGGCGCGCGCTGTCA 754  
|||||  
211 eGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeuPheA 228  
755 ACTTAAAGGACAGTCGTGGCATCAATTCGCAAAATATACAGCGCAGC 804  
|||||  
228 snLeuLysGlyGlnValValGlyIleAsnSerGlnIleTyrSerArgSer 244  
805 GCGGATTTCATGGGCATCTCTTTGCCATCCGATTGACGTTGCCATGAA 854  
|||||  
245 GlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaMetAs 261  
855 TGTCGCCGCAACAGCTGAAACACACGGCAAGTCCACGCGGACAACTGG 904  
|||||  
261 nValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGlnLeuG 278  
905 GCGTGATTATTACGAAAGTATCCTACGGTTTGGCACAGTCGTTGCTGTG 954  
|||||  
278 lyValIleIleGlnGluValSerTyrGlyLeuAlaGlnSerPheGlyLeu 294  
955 GATAAAGCC 963  
|||||  
295 AspLysAla 297

seq\_name: /cqn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y70413

seq\_documentation\_block:

ID Y70413 standard; Protein; 498 AA.

AC Y70413;

XX 03-JUL-2000 (first entry)

XX Neisseria meningitidis NMASP protein-2.

KW NMASP; non-cytosolic; antibacterial; antiinflammatory; cytotoxic;  
KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicaemia.

OS Neisseria meningitidis.

PH Key Location/Qualifiers

FT Misc-difference 8 /note= "Encoded by TTC"  
FT Misc-difference 17 /note= "Encoded by TTG"  
FT Misc-difference 22 /note= "Encoded by GAA"  
FT Misc-difference 42 /note= "Encoded by GAA"  
FT Misc-difference 58 /note= "Encoded by GCC"  
FT Misc-difference 84 /note= "Encoded by GGC"  
FT Misc-difference 134 /note= "Encoded by AAC"  
FT Misc-difference 145 /note= "Encoded by GCC"  
FT Misc-difference 194 /note= "Encoded by AAT"  
FT Misc-difference 218 /note= "Encoded by GGCATC"  
FT Misc-difference 321 /note= "Encoded by AGC"  
FT Misc-difference 340 /note= "Encoded by AGG"  
FT Misc-difference 387 /note= "Encoded by GCC"  
FT Misc-difference 396 /note= "Encoded by ACC"  
FT Misc-difference 433 /note= "Encoded by AAA"  
FT Misc-difference 451 /note= "Encoded by AAA"

FT /note= "Encoded by CAC"  
FT Misc-difference 459  
FT /note= "Encoded by AGG"  
FT Misc-difference 460  
FT /note= "Encoded by GCA"  
FT Misc-difference 461  
FT /note= "Encoded by AGT"  
FT Misc-difference 463  
FT /note= "Encoded by CGT"  
FT Misc-difference 464  
FT /note= "Encoded by CAA"  
FT Misc-difference 465..498  
FT /note= "Encoded by AGGGCAAGTCCCGCTCAA"

XX WO200012535-A2.  
XX 09-MAR-2000.  
XX 01-SEP-1999; 99WO-US19663.  
XX 01-SEP-1998; 98US-0098685.  
XX (ANTE-) ANTEX BIOLOGICS INC.  
XX Jackson WJ, Harris AM;  
XX WPI; 2000-256581/22.  
XX N-PSDB; 251538.  
XX Neisseria meningitidis MNASP polypeptide, nucleotide sequences and  
XX antibodies, useful in vaccines against infection -  
XX Claim 5; Page 70-71; 75pp; English.  
XX The present sequence is the Neisseria meningitidis MNASP protein.  
XX MNASP is a non-cytosolic protein, with antibacterial and  
XX antiinflammatory activity. It shows sequence similarity to E. coli  
XX DcpH (HtrA) protein. MNASP proteins can be used as ligands to detect  
XX antibodies elicited in response to N. meningitidis infections. Cytotoxic  
XX anti-MNASP antibodies can be used as vaccines. MNASP proteins and DNA  
XX may be used for diagnosis, therapy or prophylaxis of Neisserial  
XX infections such as, bacterial meningitis and septicaemia.

XX Sequence 498 AA;

alignment\_scores:  
Quality: 103.00 Length: 103  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y70413 ..

Align seg 1/1 to: Y70413 from: 1 to: 498

655 GTGTCCGCCAAAGCGAGAGCCTGCCAAGAAAGCTACACACCCCTTCAT 704  
218 ValSerAlaLysGlyArgSerLeuProAsnGluSerTyrThrProPheII 234  
705 CCNAAACGAGTTGCCATCAATCCGGCAATTCGGCGGCCGCGCTGTTC 754  
234 eGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeuPheA 251  
755 ACTTAAAGACAGGTCGTCCGATCAATTCGCAATATACAGCCGACG 804  
251 snLeuLysGlyGlnValValGlyIleAsnSerGlnIleTyrSerArgSer 267  
805 GGCGGATTCATGGGCAFTCTCTTTGGCATCCCGATTACGTTGCCATGAA 854  
268 GlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaMetAs 284  
855 TGTCGCGGACACTGNAANAACCCGCAAGTCCNACGGGACNACTGG 904

|||||  
284 nValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGlnLeuG 301  
905 GCGTGATTATTTCAGGAAGTATCCTACGGTTTGCACAGTCGTTCGGTCTG 954  
|||||  
301 lyValIleIleGlnGluValSerTyrGlyLeuAlaGlnSerPheGlyLeu 317  
955 GATAAAGCC 963  
|||||  
318 AspLysAla 320

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y75750

seq\_documentation\_block:

ID Y75750 standard; Protein; 499 AA.

XX Y75750;

XX 21-MAR-2000 (first entry)

XX Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2972.

XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;  
KW antibacterial; gene therapy.

XX Neisseria meningitidis.

XX WO9957280-A2.

XX 11-NOV-1999.

XX 30-APR-1999; 99WO-US09346.

XX 01-MAY-1998; 98US-0083758.

XX 31-JUL-1998; 98US-0094869.

XX 02-SEP-1998; 98US-0098994.

XX 02-SEP-1998; 98US-0099062.

XX 09-OCT-1998; 98US-0103749.

XX 09-OCT-1998; 98US-0103794.

XX 25-FEB-1999; 99US-0121528.

XX (CHIR ) CHIRON CORP.

XX (GENO-) INST GENOMIC RES.

XX Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;

XX Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;

XX Tettelin H, Venter JC;

XX WPI; 2000-062150/05.

XX N-PSDB; 254512.

XX Novel Neisserial polypeptides predicted to be useful antigens for

XX vaccines and diagnostics -

XX Claim 2; Page 1392-1393; 1453pp; English.

XX Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent  
XX novel Neisseria meningitis and N. gonorrhoeae polynucleotides and  
XX polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR  
XX primers used in the exemplification of the present invention. The  
XX polypeptides, the polynucleotides, antibodies and compositions of  
XX the invention can be used as vaccines, as diagnostic reagents, and as  
XX immunogenic compositions. The polypeptides can be used in the  
XX manufacture of medicaments for treating or preventing infection due to  
XX Neisserial bacteria (e.g. meningitis and septicaemia), to detect the  
XX presence of Neisseria bacteria, or to raise antibodies. They may also  
XX be used to screen for agonists or antagonists, which may themselves  
XX have use as antibacterial agents. The polynucleotides of the invention  
XX may also be used in gene therapy protocols.

XX Sequence 499 AA;



```

alignment_scores:
  Quality: 103.00      Length: 103
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-388-090-3 x Y5750

  Align seg 1/1 to: Y5750 from: 1 to: 499

655 GTGTCCGCCAAGGAGAGCCCTGCCCAACGAAAGCTACACACCTTCAT 704
|||||
219 ValSerAlaLysGlyArgSerLeuProAsnGluSerTyrThrProPheII 235
705 CCNAACCGAGCTTGCCATCAATCCGGGCAATTCGGGGCCCGCTGTCTCA 754
|||||
235 eGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyProLeupheA 252
755 ACTTAAAGGACAGGTCGTCGGCATCAATTCGCAATATACACGCCGAGC 804
|||||
252 snLeuLysGlyGlnValValGlyIleAsnSerGlnIleTyrSerArgSer 268
805 GCGGATTCATGGCATCTCTCTTGCATCCCGATTGACCTTGCCATGAA 854
|||||
269 GlyGlyPheMetGlyIleSerPheAlaIleProIleAspValAlaMetAs 285
855 TGTCCGCGAACAGCTGAAACACCGCGCAAGTCCACGCGGACACTGG 904
|||||
285 nValAlaGluGlnLeuLysAsnThrGlyLysValGlnArgGlyGlnLeuG 302
905 GCGTGATTATTACGGAAGTATCTACGGTTTGGCACAGTCTGTTCCGGTCTG 954
|||||
302 LyValIleIleGlnGluValSerTyrGlyLeuAlaGlnSerPheGlyLeu 318
955 GATAAAGCC 963
|||||
319 AspLysAla 321

seq_name: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y83152

seq_documentation_block:
  ID Y83152 standard; Peptide; 51 AA.
  AC Y83152;
  XX
  XX
  DT 24-JUL-2000 (first entry)
  XX
  DE NGSP peptide of Neisseria gonorrhoeae.
  XX
  KW NGSP; polypeptide; peptide; vaccine; immune response; antibody;
  cellular matrix; adherence domain; ligand; detection; diagnosis;
  screening; probe; primer; prophylaxis; therapy.
  XX
  OS Neisseria gonorrhoeae.
  XX
  PN WQ200012133-A1.
  XX
  PD 09-MAR-2000.
  XX
  PF 01-SEP-1999; 99WO-US20070.
  XX
  PR 01-SEP-1998; 98US-0098685.
  XX
  PA (ANTE-) ANTEX BIOLOGICS INC.
  XX
  PI Jackson WJ, Harris AM;
  XX
  DR WPI: 2000-237782/20.
  DR N-PSDB; 293416.
  XX
  PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from

```

```

PT
XX
PS
XX
XX
CC Claim 5; Page 64-65; 68pp; English.
CC
CC The NGSP polypeptide of N. gonorrhoeae has conserved Arg-Gly-Asp and
Arg-Gly-Asn motifs near the C-terminus which function as adherence
domains for extracellular matrix proteins. Using the NGSP polypeptide
as a vaccine produces antibodies which inhibit binding of N.
gonorrhoeae to the host's cellular matrix reducing attachment and/or
subsequent invasion. The NGSP polypeptide and its peptide fragments
can be used to immunise an animal and produce an immune response.
They can also be used as ligands to detect antibodies elicited in
response to Neisseria infections and also as antigens or immunogens
for inducing Neisseria-specific antibodies which are useful in
immunoassays to detect Neisseria in biological specimens. Nucleotides
encoding NGSP or its fragments can be used as probes to identify
Neisseria in biological specimens by hybridization or polymerase
chain reaction amplification. The NGSP polypeptide can also be used
in screening assays to identify agents and compounds which useful as
diagnostic, prophylactic or therapeutic agents against Neisseria
infection.
XX
SQ Sequence 51 AA;

alignment_scores:
  Quality: 51.00      Length: 51
  Ratio: 1.000        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-388-090-3 x Y83152

  Align seg 1/1 to: Y83152 from: 1 to: 51

1 GTGTTCAAAAATACCAATCTTCCTTTGGCGCACTGTGCGCCCTT 50
|||||
1 ValPheLysLysTyrGlnTyrPheAlaLeuAlaLeuLysAlaAlaLe 17
51 GCTGGCAGGTCGGAAAGGCGAGGAGCTTTTCGGTGGCGGACAAAAG 100
|||||
17 uLeuAlaGlyCysGluLysAlaGlySerPhePheGlyAlaAspLysLysG 34
101 AAGCATCTTCGTAGAACGCATCGACACACCAAGACGACGCGCAGTGC 150
|||||
34 luAlaSerPheValGluArgIleGluHisThrLysAspAspGlySerVal 50
151 AGT 153
|||
51 Ser 51

seq_name: /cgn2_2/gcgdata/geneseq/geneseq/AA2000.DAT:Y52997

seq_documentation_block:
  ID Y52997 standard; Protein; 129 AA.
  XX
  AC Y52997;
  XX
  DT 21-FEB-2000 (first entry)
  XX
  DE Neisseria meningitidis variable BASB013-V protein sequence.
  XX
  KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
  antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
  invasive bacterial disease; antibacterial.
  XX
  OS Neisseria meningitidis.
  XX
  PN WO9955872-A1.
  XX
  PD 04-NOV-1999.
  XX

```

PF 20-APR-1999; 99WO-EP02765.  
XX  
PR 23-APR-1998; 98GB-0008734.  
XX  
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX  
PI Ruelle J;  
XX  
DR WPI: 2000-052809/04.  
DR N-PSDB; 233309.  
XX  
XX Novel polynucleotides and polypeptides from Neisseria meningitis used  
PT to prepare vaccines against bacterial infections -  
XX  
XX Example 2; Page 83-84; 94pp; English.  
XX  
CC The present sequence is a variable BASB013-V polypeptide isolated  
CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides  
CC may be employed as research reagents and material for the discovery of  
CC treatments and diagnostics for diseases, particularly human diseases.  
CC They can be used for diagnosis of disease, staging of disease, or  
CC determining response of an infectious organism to drugs. The  
CC polynucleotides may be used as a source for hybridisation probes, and  
CC for screening of genetic mutations, serotype, organism or strain  
CC identification, identification of mutation in BASB013 sequences, and as  
CC components of arrays which are useful for diagnostic and prognostic  
CC purposes. The polypeptides can be used to produce antibodies. The  
CC polypeptides can also be used in vaccine formulations, and to identify  
CC agonists and antagonists. The polypeptides, antibodies, agonists and  
CC antagonists (which are bacteriostatic) are used for the treatment and  
CC prevention of diseases such as upper respiratory tract infection,  
CC invasive bacterial diseases such as bacteraemia and meningitis, and for  
CC the development and screening of antibacterial drugs. They are also used  
CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins  
CC on in-dwelling devices, or to extracellular proteins on wounds, and to  
CC thus prevent tissue damage and/or block the normal progression of  
CC pathogenesis in infections initiated other than by the implantation of  
CC in-dwelling devices or by other surgical techniques.  
XX  
SQ Sequence 129 AA;

alignment\_scores:  
Quality: 36.00 Length: 36  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y52997 ..

Align seg 1/1 to: Y52997 from: 1 to: 129

1192 GCGCATCATCCAAAACAGATGAAGCCCTACACCGAAGCAGCATCCGG 1241

|||||  
28 GlyAlaSerSerLysThrAspGluAlaProTyrThrGluGlnGlnSerGI 44

1242 TAGGTTCTCGGTGGAATCCCGAGGCATTACCCCTTCAGACACATACCCGACA 1291

|||||  
44 yThrPheSerValGluSerAlaGlyIleThrLeuGlnThrHisThrAsps 61

1292 GCAGCGGC 1299

|||||  
61 erSerGly 63

seq\_name: /cgn2\_2/gcgdata/geneseq/geneseqp/AA2000.DAT:Y70415

seq\_documentation\_block:

ID\_ Y70415 standard; Protein: 51 AA.

XX Y70415;

XX 03-JUL-2000 (first entry)

XX

DE Neisseria meningitidis NMA SP protein N-terminal fragment.  
XX  
KW NMA SP: non-cytosolic; antibacterial; antiinflammatory; cytolotoxic;  
KW anti-NMA SP antibody; vaccine; diagnosis; therapy; prophylaxis;  
KW Neisserial infection; meningitidis; septicemia.  
XX  
OS Neisseria meningitidis.  
XX  
PN WO200012535-A2.  
XX  
PD 09-MAR-2000.  
XX  
PF 01-SEP-1999; 99WO-US19663.  
XX  
PR 01-SEP-1998; 98US-0098685.  
XX  
PA (ANTE-) ANTEX BIOLOGICS INC.  
XX  
PI Jackson WJ, Harris AM;  
XX  
DR WPI: 2000-256581/22.  
DR N-PSDB; 251542.  
XX  
XX Neisseria meningitidis NMA SP polypeptide, nucleotide sequences and  
PT antibodies, useful in vaccines against infection -  
XX  
PS Disclosure; Page 74; 75pp; English.  
XX  
CC The present sequence is the N-terminal fragment of Neisseria  
CC meningitidis NMA SP protein. NMA SP is a non-cytosolic protein, with  
CC antibacterial and antiinflammatory activity. It shows sequence similarity  
CC to E. coli DegP (HtrA) protein. NMA SP proteins can be used as ligands to  
CC detect antibodies elicited in response to N. meningitidis infections.  
CC Cytotoxic anti-NMA SP antibodies can be used as vaccines. NMA SP proteins  
CC and DNA may be used for diagnosis, therapy or prophylaxis of Neisserial  
CC infections such as, bacterial meningitidis and septicemia.  
XX  
SQ Sequence 51 AA;

alignment\_scores:

Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Y70415 ..

Align seg 1/1 to: Y70415 from: 1 to: 51

67 AAGCAGCGCAGCTTTTCGGTCCGACAAAAGAACGATCCTTCGTAGA 116

|||||  
23 LysAlaGlySerPheGlyAlaAspLysGluAlaSerPheValGI 39

117 ACGCATC 123

|||||

39 uArgIle 41





OM of: US-09-388-090-3 to: Issued\_Patents\_AH:\* out\_format : pfs

Date: Mar 22, 2001 12:05 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODET=frame\*\_n2p.model -DEV=xlpl  
-O=/cgn2\_1/USPTO\_spool/US09388090/runat\_20032001\_092950\_25795/app\_query.fasta\_1.1474  
-DB=Issued\_Patents\_AA -OFMT=Eastan -SUFFIX=oligo7.rai  
-GAPOP=4.500 -GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000  
-LOOPEXT=0.000 -QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=60.000  
-XGAPEXT=60.000 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000  
-YGAPEXT=60.000 -DELOP=6.000 -DELEXT=7.000 -START=1  
-MATRIX=oligo -TRANS=human40.cdi -LIST=1000 -DOCALIGN=200  
-THR\_SCORE=quality -THR\_MIN=7 -ALIGN=15 -MODE=LOCAL -OUTEMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=200000000  
-USER=US09388090@cgn1\_1\_19 -NCPU=6 -ICPU=3 -LONGLOG -NO\_XLPHY  
-WAIT -THREADS=1

Search information block:

Query: US-09-388-090-3

Query length: 1395

Database: Issued\_Patents\_AH:\*

Database sequences: 174772

Database length: 17957048

Search time (sec): 23.850000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd Orig	zScore	Len	Documentation
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-09-142-469-3	8.00	133.43	16.49	27
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-09-142-469-2	8.00	117.13	12.13	297
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-245-294-8	8.00	113.55	11.33	503
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-479-818-8	8.00	113.55	11.33	503
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-307-279A-8	8.00	113.55	11.33	503
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-351-2	8.00	112.90	11.20	553
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-445-586-7	8.00	112.89	11.20	554
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-471-341-7	8.00	112.87	11.19	556
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-669-721-18	7.00	122.14	172.16	11
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-320-373-80	7.00	119.59	164.08	16
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-320-373-80	7.00	119.59	164.08	16
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-637-759B-412	7.00	119.59	164.08	16
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-871-355A-412	7.00	119.59	164.08	16
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-297-395-7	7.00	118.43	160.50	19
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-464-318-7	7.00	116.04	153.43	27
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-471-341-7	7.00	116.04	153.43	27
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-461-566-7	7.00	116.04	153.43	27
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-297-395-10	7.00	107.56	130.74	94
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-466-860-7	7.00	106.31	127.69	113
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-906-769-163	7.00	106.31	127.69	113
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-472-040A-7	7.00	106.31	127.69	113
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-906-616-163	7.00	106.31	127.69	113
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-639-075A-163	7.00	106.31	127.69	113
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-2	7.00	103.41	120.90	173
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-487-550-6	7.00	101.22	115.99	239
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-923-454A-4	7.00	99.17	111.60	323
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-773-870-6	7.00	99.13	111.51	325
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-468-853-6	7.00	99.02	111.29	330
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-468-855-6	7.00	99.02	111.29	330
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-310-357-6	7.00	99.02	111.29	330
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-468-852-6	7.00	99.02	111.29	330
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-468-852-6	7.00	99.02	111.29	330
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-27	7.00	98.12	109.41	377
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-09-108-020-53	7.00	97.87	108.90	391
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-2	7.00	97.87	108.90	391
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-742-440A-8	7.00	97.82	108.79	394
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-097-938-2	7.00	97.80	108.75	395

/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-097-938-5	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-476-000-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-476-000-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-472-840-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-472-840-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-474-410-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-474-410-2	7.00	97.80	108.75	395
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-476-000-63	7.00	97.77	108.68	397
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-472-840-63	7.00	97.77	108.68	397
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-63	7.00	97.77	108.68	397
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-63	7.00	97.77	108.68	397
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-097-938-6	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-097-938-6	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-476-000-6	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-476-000-6	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-472-840-4	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-472-840-4	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-6	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-6	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-474-410-4	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-474-410-4	7.00	97.75	108.65	398
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-476-000-61	7.00	97.73	108.61	399
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-472-840-61	7.00	97.73	108.61	399
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-61	7.00	97.73	108.61	399
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-476-978-61	7.00	97.73	108.61	399
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-8	7.00	96.80	106.71	458
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-8	7.00	96.80	106.71	458
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-18	7.00	96.48	106.07	480
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-923-454A-18	7.00	96.48	106.07	480
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-811-897A-23	7.00	95.87	104.86	525
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-811-897A-23	7.00	95.87	104.86	525
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-855-213-23	7.00	95.75	104.63	534
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-855-213-23	7.00	95.75	104.63	534
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-669-721-3	7.00	95.60	104.33	546
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-669-721-3	7.00	95.60	104.33	546
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-811-897A-27	7.00	95.53	104.18	552
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-811-897A-27	7.00	95.53	104.18	552
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-811-897A-25	7.00	95.51	104.16	553
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-811-897A-25	7.00	95.51	104.16	553
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-855-213-25	7.00	95.51	104.16	553
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-855-213-25	7.00	95.51	104.16	553
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-892-770-6	7.00	95.37	103.87	565
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-892-770-6	7.00	95.37	103.87	565
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-606-288-7	7.00	94.95	103.05	601
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-606-288-7	7.00	94.95	103.05	601
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-09-347-483-7	7.00	94.95	103.05	601
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-09-347-483-7	7.00	94.95	103.05	601
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-985-950-16	7.00	94.79	102.75	615
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-985-950-16	7.00	94.79	102.75	615
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-985-950-18	7.00	94.79	102.75	615
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-985-950-18	7.00	94.79	102.75	615
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-559-397A-31	7.00	93.97	101.17	694
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-559-397A-31	7.00	93.97	101.17	694
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-928-692-17	7.00	92.56	98.51	854
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-928-692-17	7.00	92.56	98.51	854
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-554-613C-1	7.00	92.36	98.15	879
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-554-613C-1	7.00	92.36	98.15	879
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-968-752B-4	7.00	91.25	96.10	1036
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-968-752B-4	7.00	91.25	96.10	1036
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-768-147B-2	7.00	91.00	95.66	1074
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-768-147B-2	7.00	91.00	95.66	1074
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-015-986A-3	7.00	89.00	92.11	1442
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-015-986A-3	7.00	89.00	92.11	1442
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-446-363-3	7.00	88.99	92.09	1445
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-446-363-3	7.00	88.99	92.09	1445
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-446-363A-2	7.00	88.99	92.09	1445
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-446-363A-2	7.00	88.99	92.09	1445
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-494-168-2	7.00	87.91	90.23	1694
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-08-494-168-2	7.00	87.91	90.23	1694
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-459-568-4	7.00	87.81	90.06	1719
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-459-568-4	7.00	87.81	90.06	1719
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-516-859A-4	7.00	87.81	90.06	1719
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-08-516-859A-4	7.00	87.81	90.06	1719
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-09-335-409-4	7.00	87.37	89.32	1832
/cgn2_6/ptodata/2/iaa/6/COMB.ppt:US-09-335-409-4	7.00	87.37	89.32	1832
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-07-745-208A-7	7.00	86.89	88.51	1968
/cgn2_6/ptodata/2/iaa/5A/COMB.ppt:US-07-745-208A-7	7.00	86.89	88.51	1968
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-455-543A-45	7.00	86.89	88.51	1968
/cgn2_6/ptodata/2/iaa/5B/COMB.ppt:US-08-455-543A-45	7.00	86.89	88.51	1968

```
/cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-311-363-7 + 7.00 86.89 88.51 1968
/cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-276-967-2 - 7.00 85.33 85.94 2476
/cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-149-097D-35 + 7.00 85.24 85.79 2509
/cgn2_6/ptodata/2/iaa/6_COMB.pap:US-09-036-987A-4 - 7.00 83.65 83.26 3170
/cgn2_6/ptodata/2/iaa/5A_COMB.pap:US-08-469-486-52 - 7.00 81.20 79.09 4544
/cgn2_6/ptodata/2/iaa/5B_COMB.pap:US-08-469-658-52 - 7.00 81.20 79.09 4544

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pap:US-09-142-469-3

seq_documentation_block:
: Sequence 3, Application US/09142469
: Patent No. 6140094
: GENERAL INFORMATION:
: APPLICANT: LOFFLER, Fridolin
: APPLICANT: JUNGSCHAFFER, Gerald
: APPLICANT: KHANH, Quoc Nguyen
: APPLICANT: SCHUSTER, Erwin
: APPLICANT: SPROSSLER, Bruno
: APPLICANT: WOLF, Sabine
: TITLE OF INVENTION: PROTEIN WITH PHOSPHOLIPASE ACTIVITY
: FILE REFERENCE: 015200-056
: CURRENT APPLICATION NUMBER: US/09/142,469
: EARLIER FILING DATE: 1998-09-08
: EARLIER APPLICATION NUMBER: PCT/EP98/00081
: EARLIER FILING DATE: 1998-01-08
: EARLIER APPLICATION NUMBER: DE 197 01 348.1
: EARLIER FILING DATE: 1997-01-16
: NUMBER OF SEQ ID NOS: 8
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 27
: TYPE: PRT
: ORGANISM: Aspergillus lysophospholipase
: US-09-142-469-3

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-09-142-469-3 ..
Align seg 1/1 to: US-09-142-469-3 from: 1 to: 27

25 GCTTTGGCGGCGACTGTGTCGCCGC 48
|||||
12 AlaLeuAlaAlaLeuCysAlaAla 19

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pap:US-09-142-469-2

seq_documentation_block:
: Sequence 8, Application US/08245294
: Patent No. 5644047
: GENERAL INFORMATION:
: APPLICANT: Anderson, Burt E.
: APPLICANT: Regnery, Russell L.
: TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
: TITLE OF INVENTION: DIAGNOSING
: TITLE OF INVENTION: ROCHALIMAEAE HENSELAE AND ROCHALIMAEAE QUINTANA
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C.
: STREET: 127 Peachtree Street, Suite 1200
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/245,294
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Spratt, Gwendolyn D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 1414.612
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404/688-0770
: TELEFAX: 404/688-9880
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 503 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-245-294-8

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-245-294-8 ..
```

```
Align seq 1/1 to: US-08-245-294-8 from: 1 to: 503

640 GTGACCGCGGCATCGTGTCGCC 663
|||||
215 ValThrAlaGlyIleValSerAla 222

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-474-499-8

seq_documentation_block:
; Sequence 8, Application US/08474499
; Patent No. 5693776
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
; TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
; TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/474,499
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.612
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-474-499-8

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-474-499-8 ..
Align seq 1/1 to: US-08-474-499-8 from: 1 to: 503

640 GTGACCGCGGCATCGTGTCGCC 663
|||||
215 ValThrAlaGlyIleValSerAla 222

seq_name: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:US-08-307-279A-8

seq_documentation_block:
; Sequence 8, Application US/08307279A
; Patent No. 5736347
; GENERAL INFORMATION:
```

```
; APPLICANT: Anderson, Burt E.
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
; TITLE OF INVENTION: Henselae and Rochalimaea Quintana Infection
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,279A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-307-279A-8

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-307-279A-8 ..
Align seq 1/1 to: US-08-307-279A-8 from: 1 to: 503

640 GTGACCGCGGCATCGTGTCGCC 663
|||||
215 ValThrAlaGlyIleValSerAla 222

seq_name: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:PCT-US95-06211-8

seq_documentation_block:
; Sequence 8, Application PC/TUS9506211
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
```

3009535147

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06211
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/245,294
; FILING DATE: 18 MAY 1994
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.6121
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06211-8

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-388-090-3 x PCT-US95-06211-8 ..
  Align seg 1/1 to: PCT-US95-06211-8 from: 1 to: 503

  640 GTGACGCGCGCATCGTGCCGCC 663
  |||||
  215 ValThrAlaGlyIleValSerAla 222

seq_name: /cgn2_6/ptodata/2/1aa/6_COMB.pep:us-09-083-351-2

seq_documentation_block:
; Sequence 2, Application US/09083351
; Patent No. 6087107
; GENERAL INFORMATION:
; APPLICANT: Sheffield, Val C.
; APPLICANT: Alward, Wallace L.M.
; APPLICANT: Stone, Edwin M.
; APPLICANT: Nishimura, Darryl
; APPLICANT: Patil, Shiva
; TITLE OF INVENTION: THERAPEUTICS AND DIAGNOSTICS FOR
; TITLE OF INVENTION: CONGENITAL HEART DISEASE BASED ON A NOVEL HUMAN
; TITLE OF INVENTION: TRANSCRIPTION FACTOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/083,351
; FILING DATE: 22-MAY-1998
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold, Beth E.
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: UIA-029.02
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000

; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 553 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-083-351-2

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-388-090-3 x US-09-083-351-2 ..
  Align seg 1/1 to: US-09-083-351-2 from: 1 to: 553

  599 GGGTCGCTGCCATCGCGCGCCT 622
  |||||
  275 GlySerLeuProSerAlaArgPro 282

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-445-586-7

seq_documentation_block:
; Sequence 7, Application US/08445586
; Patent No. 5627050
; GENERAL INFORMATION:
; APPLICANT: Takeshita, Sunao
; APPLICANT: Ito, Toshimi
; APPLICANT: Otawara-Hamamoto, Yoko
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Sulfatase-Like Protein and
; TITLE OF INVENTION: Process for its Production
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,586
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/111,887
; FILING DATE: 26-AUG-1993
; APPLICATION NUMBER: JP 230030/92
; FILING DATE: 28-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324034/92
; FILING DATE: 03-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1322-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 554 amino acids
; TYPE: amino acid
```



```
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-445-586-7

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-445-586-7 ..
Align seg 1/1 to: US-08-445-586-7 from: 1 to: 554

217 GCAGCCCCCGCCGCGCACCCAA 240
|||||
18 AlaAlaProAlaProArgThrGln 25

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-445-586-2

seq_documentation_block:
; Sequence 2, Application US/08445586
; Patent No. 5627050
; GENERAL INFORMATION:
; APPLICANT: Takeshita, Sunao
; APPLICANT: Ito, Toshimi
; APPLICANT: Otawara-Hamamoto, Yoko
; APPLICANT: Amann, Egon
; TITLE OF INVENTION: Bone-Related Sulfatase-Like Protein and
; TITLE OF INVENTION: Process for its Production
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSER: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,586
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/111,887
; FILING DATE: 26-AUG-1993
; APPLICATION NUMBER: JP 230030/92
; FILING DATE: 28-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 324034/92
; FILING DATE: 03-DEC-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Forman, David S.
; REGISTRATION NUMBER: 33,694
; REFERENCE/DOCKET NUMBER: 02481.1322-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 556 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-445-586-2
```

```
alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-445-586-2 ..
Align seg 1/1 to: US-08-445-586-2 from: 1 to: 556

217 GCAGCCCCCGCCGCGCACCCAA 240
|||||
18 AlaAlaProAlaProArgThrGln 25

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-669-721-18

seq_documentation_block:
; Sequence 18, Application US/08669721
; Patent No. 5834236
; GENERAL INFORMATION:
; APPLICANT: Lamb et al., Christopher J.
; TITLE OF INVENTION: NOVEL TRANSCRIPTION ENHANCER ELEMENT AND
; TITLE OF INVENTION: TRANSCRIPTION FACTOR AND METHODS OF USE THEREFOR
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/669,721
; FILING DATE: 27-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellison, Eldora L.
; REGISTRATION NUMBER: 39,967
; REFERENCE/DOCKET NUMBER: 07251/014001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-669-721-18

alignment_scores:
  Quality: 7.00      Length: 7
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-669-721-18 ..
Align seg 1/1 to: US-08-669-721-18 from: 1 to: 11

180 GGTTCAAAGCGAAGCCCGC 200
|||||
1 GlySerLysArgArgProGly 7

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-320-373-79
```

```

seq_documentation_block:
; Sequence 79, Application US/08320373
; Patent No. 5559025
; GENERAL INFORMATION:
; APPLICANT: Ahorn, Horst
; APPLICANT: Maurer-Poggy, Ingrid
; APPLICANT: Sommergruber, Wolfgang
; APPLICANT: Zophel, Andreas
; APPLICANT: Blaas, Dieter
; APPLICANT: Kuchler, Ernst
; APPLICANT: Liebig, Hans-Dieter
; APPLICANT: Skern, Timothy
; TITLE OF INVENTION: Expression of Mature Proteinase 2A, the
; TITLE OF INVENTION: Partial Purification Thereof and Preparation of Substrates
; TITLE OF INVENTION: Having an Inhibitory Effect
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320.373
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,619
; FILING DATE: 06-NOV-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-320-373-79

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3/rev x US-08-320-373-79 ..
Align seg 1/1 to: US-08-320-373-79 from: 1 to: 16

209 TTGAGCTGCGGGCCTTCG 189
|||||
5 LeuThrThrAlaGlyProSer 11

seq_name: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:US-08-320-373-80

seq_documentation_block:
; Sequence 80, Application US/08320373
; Patent No. 5559025
; GENERAL INFORMATION:
; APPLICANT: Ahorn, Horst
; APPLICANT: Maurer-Poggy, Ingrid
; APPLICANT: Sommergruber, Wolfgang
; APPLICANT: Zophel, Andreas
; APPLICANT: Blaas, Dieter
; APPLICANT: Kuchler, Ernst
; TITLE OF INVENTION: Expression of Mature Proteinase 2A, the
; TITLE OF INVENTION: Partial Purification Thereof and Preparation of Substrates
; TITLE OF INVENTION: Having an Inhibitory Effect
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320.373
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,619
; FILING DATE: 06-NOV-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-320-373-79

```

```

; APPLICANT: Liebig, Hans-Dieter
; APPLICANT: Skern, Timothy
; TITLE OF INVENTION: Expression of Mature Proteinase 2A, the
; TITLE OF INVENTION: Partial Purification Thereof and Preparation of Substrates
; TITLE OF INVENTION: Having an Inhibitory Effect
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/320.373
; FILING DATE: 11-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,619
; FILING DATE: 06-NOV-1992
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 466-0800
; TELEFAX: (202) 833-8716
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-320-373-80

alignment_scores:
Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3/rev x US-08-320-373-80 ..
Align seg 1/1 to: US-08-320-373-80 from: 1 to: 16

209 TTGAGCTGCGGGCCTTCG 189
|||||
5 LeuThrThrAlaGlyProSer 11

seq_name: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:US-08-637-759B-412

seq_documentation_block:
; Sequence 412, Application US/08637759B
; Patent No. 5876931
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Rabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,759B
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: RPMS 101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 412:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-637-759B-412

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-637-759B-412 ..
Align seg 1/1 to: US-08-637-759B-412 from: 1 to: 16

790 ATATACAGCGCGCGCGGA 810
|||||
2 IleTyrSerArgSerGlyGly 8

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-871-355A-412

seq_documentation_block:
; Sequence 412, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/871,355A
; FILING DATE: 09-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB95/02875
; FILING DATE: 11-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284

```

```

; REFERENCE/DOCKET NUMBER: RPMS 101 CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 412:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; US-08-871-355A-412

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x US-08-871-355A-412 ..
Align seg 1/1 to: US-08-871-355A-412 from: 1 to: 16

790 ATATACAGCGCGCGCGGA 810
|||||
2 IleTyrSerArgSerGlyGly 8

seq_name: /cgn2_6/ptodata/2/iaa/6_COMB.pep:US-08-297-395-7

seq_documentation_block:
; Sequence 7, Application US/08297395A
; Patent No. 6039947
; GENERAL INFORMATION:
; APPLICANT: Howard L. Weiner
; APPLICANT: David A. Hafler
; TITLE OF INVENTION: PEPTIDES DERIVED FROM IMMUNODOMINANT
; TITLE OF INVENTION: EPITOPES OF MYELIN BASIC PROTEIN
; FILE REFERENCE: 1010/05723US3
; CURRENT APPLICATION NUMBER: US/08/297,395A
; CURRENT FILING DATE: 1994-08-11
; EARLIER APPLICATION NUMBER: 08/059,189
; EARLIER FILING DATE: 1993-05-06
; EARLIER APPLICATION NUMBER: 07/502,559
; EARLIER FILING DATE: 1990-03-30
; EARLIER APPLICATION NUMBER: PCT/US88/02139
; EARLIER FILING DATE: 1988-06-24
; EARLIER APPLICATION NUMBER: 07/065,734
; EARLIER FILING DATE: 1987-06-24
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 19
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-297-395-7

alignment_scores:
  Quality: 7.00 Length: 7
  Ratio: 1.000 Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3/rev x US-08-297-395-7 ..
Align seg 1/1 to: US-08-297-395-7 from: 1 to: 19

534 GATTTTCAGAAAGGCGCATC 514
|||||
8 AspPheGlnLysGlyAspIle 14

```



OM of: US-09-388-090-3 to: PIR\_66:\* out\_format : pfs

Date: Mar 22, 2001 12:05 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09388090/runat\_20032001\_092950\_25808/app\_query.fasta\_1.1474  
-DB=PIR\_66 -OFMT=fastan -SUPFI=oligo7.rpr -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPEXT=60.000 -GAPEXT=60.000  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -GAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=1000 -DOCALIGN=200 -THR\_SCORE=quality  
-THR\_MTN=7 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09388090@cgn\_1\_171 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-388-090-3

Query length: 1395

Database: PIR\_66.\*

Database sequences: 195891

Database length: 67900655

Search time (sec): 40.680000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score_list:	Sequence	Strd	Orig	zScore	EScore	Len	Documentation
+	pir2:B81914	+	127.00	2307.34	2.8e-121	499	probable periplasmic serine protease
+	pir2:F83550	+	19.00	319.86	1.5e-10	474	serine proteinase MucD precursor
+	pir2:F70048	+	11.00	177.56	0.0267	224	serine proteinase Do homolog yv
+	pir2:S74643	+	11.00	173.83	0.0245	394	proteinase hhoA (EC 3.4.21.1)
+	pir2:B71284	+	11.00	173.76	0.0245	398	probable periplasmic serine protease
+	pir2:S75445	+	11.00	173.47	0.0243	416	proteinase hhoB (EC 3.4.21.1)
+	pir2:S75357	+	11.00	173.09	0.0241	441	probable periplasmic serine protease
+	pir2:A69644	+	11.00	172.97	0.0240	449	serine proteinase Do, heat-shock
+	pir2:S77538	+	11.00	172.92	0.0240	452	serine proteinase (EC 3.4.21.1)
+	pir1:B74326	+	11.00	172.91	0.0240	453	periplasmic serine proteinase
+	pir2:F72359	+	11.00	172.82	0.0239	459	periplasmic serine proteinase
+	pir2:G81528	+	11.00	172.42	0.0237	488	serine proteinase, HtrA/DegQ/DegP
+	pir2:B72011	+	11.00	172.30	0.0236	497	probable do serine proteinase
+	pir2:H71728	+	11.00	172.30	0.0236	497	probable do serine proteinase
+	pir2:T35460	-	9.00	138.93	2.88	295	integral membrane protein - Str
+	pir2:T35287	+	9.00	137.58	2.79	362	probable secreted proteinase
+	pir2:B83089	+	9.00	137.10	2.76	400	AlgW protein PA4446 [imported]
+	pir2:B70092	+	9.00	136.92	2.75	400	serine proteinase Do homolog yv
+	pir2:T45448	+	9.00	136.11	2.69	452	probable serine proteinase (EC
+	pir2:C70821	+	9.00	135.94	2.68	464	probable serine proteinase Rv09
+	pir1:I40059	+	9.00	135.80	2.68	474	htrA-like protein - Brucella ab
+	pir2:A82581	+	9.00	135.26	2.64	514	periplasmic proteinase XF2241
+	pir2:T02457	+	9.00	133.04	2.51	720	hypothetical protein F4118.12
+	pir2:T22203	+	9.00	132.42	2.47	790	transcription factor Vp1 - comm
+	pir2:E64040	+	9.00	132.16	2.46	822	hypothetical protein L1231.2
+	pir2:T02824	-	8.00	121.70	31.32	247	hypothetical protein H11701
+	pir2:T35304	+	8.00	120.82	30.69	282	probable peptidase - Streptomy
+	pir2:T72250	+	8.00	120.46	30.43	298	hypothetical protein F45H1.1
+	pir2:T71023	+	8.00	120.41	30.40	300	probable transcription initiation
+	pir2:S34116	+	8.00	120.41	30.40	300	transcription factor IIB - Pyro
+	pir2:T46883	+	8.00	120.41	30.40	300	transcription factor IIB [valid
+	pir2:E75110	+	8.00	120.41	30.40	300	transcription initiation factor
+	pir2:D72693	+	8.00	120.18	30.24	311	hypothetical protein APE0964
+	pir2:E71726	+	8.00	119.23	29.58	359	hypothetical protein RPI61 - Ri
+	pir1:G49964	+	8.00	119.10	29.49	366	photosynthetic reaction center
+	pir2:750891	+	8.00	119.10	29.49	366	cytochrome subunit of photosynt
+	pir2:T48361	-	8.00	118.05	28.79	429	gene GEG-154 protein - mouse
+	pir1:C64647	+	8.00	117.84	28.65	443	serine proteinase (EC 3.4.21.1)

+	pir2:B64160	8.00	117.55	28.45	463	hypothetical protein HI0852
+	pir2:F81329	8.00	117.42	28.37	472	serine proteinase (proteins)
+	pir2:H71936	8.00	117.36	28.33	476	proteinase DO - Helicobacter
+	pir2:D82826	8.00	117.30	28.29	481	heat shock protein Xf0285 [i
-	pir1:SVBSES	8.00	117.19	28.22	489	glutamate--trNA ligase (EC f
-	pir2:T36336	8.00	116.08	27.51	578	probable membrane protein -
+	pir1:F64397	8.00	115.08	26.87	673	transcription initiation fac
+	pir2:S69074	8.00	114.49	26.51	736	hypothetical protein YPR090
+	pir2:S82821	8.00	113.70	26.03	829	pyrroloquinoline quinone syn
+	pir2:T47784	8.00	111.76	24.89	1112	hypothetical protein F17J16
+	pir2:S69048	8.00	110.85	24.37	1376	probable membrane protein y
-	pir2:T43060	8.00	110.59	24.23	1328	agrin - electric ray (Disco
-	pir2:T47492	8.00	110.41	24.13	1363	copla-like polyprotein - Ar
+	pir2:T34869	8.00	109.72	23.75	1514	glutamate synthase (NADPH)
+	pir2:E44906	7.00	116.20	447.22	35	L1 protein - human papillomav
+	pir2:F82459	7.00	115.16	436.55	41	hypothetical protein VCA0437
+	pir2:S04941	7.00	114.54	430.39	45	protamine phi-3.1 - Californ
+	pir2:S10544	7.00	114.54	430.39	45	protamine phi-3.2 - Californ
+	pir2:S10545	7.00	114.54	430.39	45	protamine phi-3.3 - Californ
+	pir2:B82466	7.00	112.75	412.96	59	hypothetical protein VCA0381
-	pir2:C82461	7.00	112.75	412.96	59	hypothetical protein VCA0412
-	pir2:D82459	7.00	112.75	412.96	59	hypothetical protein VCA0435
-	pir2:B82465	7.00	112.64	411.90	60	hypothetical protein VCA0407
-	pir2:E82025	7.00	111.53	401.45	71	hypothetical protein NMA0301
-	pir2:T51537	7.00	111.53	401.45	71	hypothetical protein NMA0301
+	pir2:C81003	7.00	110.74	394.21	80	hypothetical protein T20K14.2
+	pir2:S52779	7.00	110.66	393.46	81	hypothetical protein NMB2130
+	pir2:E75417	7.00	110.19	389.19	87	T-cell receptor beta-chain V-
+	pir2:S78005	7.00	109.82	385.89	92	cuticle structural protein PC
-	pir2:H83378	7.00	109.61	384.00	95	hypothetical protein PA2143
+	pir2:JC1150	7.00	109.14	379.86	102	hypothetical protein, 11.2K
+	pir2:G64708	7.00	108.76	376.56	108	hypothetical protein HP1511
+	pir1:CCAG2	7.00	108.58	374.99	111	cytochrome c2 - Agrobacteriu
+	pir2:S10918	7.00	108.52	374.47	112	hypothetical 12K protein - E
+	pir2:A32578	7.00	108.46	373.96	113	T-cell receptor beta chain p
+	pir2:B83575	7.00	108.46	373.96	113	hypothetical protein PA0554
+	pir2:T44881	7.00	108.29	372.47	116	hypothetical protein MCB188
+	pir2:A82259	7.00	108.23	371.98	117	hypothetical protein VC0942
+	pir2:T36232	7.00	108.12	371.02	119	hypothetical protein APE2186
-	pir2:D72506	7.00	107.95	369.62	122	probable DNA binding protein
+	pir2:A83400	7.00	107.85	368.70	124	hypothetical protein APE2026
+	pir2:T72478	7.00	107.53	366.05	130	hypothetical protein APE1969
+	pir2:H05440	7.00	107.48	365.62	131	hypothetical protein APE2472
+	pir2:S67869	7.00	107.38	364.78	133	hypothetical protein F7K2.40
+	pir2:T36232	7.00	107.19	363.13	137	T cell receptor Mbl1 beta ch
-	pir2:S74898	7.00	107.19	363.13	137	hypothetical protein ybdD -
+	pir2:F83388	7.00	107.04	361.93	140	hypothetical protein APE0742
+	pir2:S83391	7.00	106.95	361.15	142	hypothetical protein sll1960
+	pir2:B83544	7.00	106.86	360.38	144	T-cell receptor beta chain p
+	pir2:S83291	7.00	106.68	358.88	148	conserved hypothetical prote
-	pir2:S65600	7.00	106.68	358.88	148	conserved hypothetical prote
+	pir2:S53659	7.00	106.63	358.51	149	hypothetical 16.4K protein
+	pir2:T29164	7.00	106.46	357.06	153	hydrogenase-related protein
+	pir2:JEO256	7.00	106.33	356.00	155	trappin-10 - collapsed peccar
+	pir2:T07174	7.00	106.33	356.00	156	tuberculosis-related protein
+	pir2:S65156	7.00	106.29	355.66	157	hypothetical 17.5 kd protein
+	pir2:A83474	7.00	106.04	353.63	163	hypothetical protein PA1381
-	pir2:S31894	7.00	105.76	351.36	170	hypothetical protein y - Myx
-	pir2:S31897	7.00	105.72	351.05	171	conserved hypothetical prote
-	pir1:NNNCCV	7.00	105.61	350.12	174	nonstructural protein C - ca
+	pir2:C71524	7.00	105.46	348.91	178	hypothetical protein CT358 -
+	pir2:T43451	7.00	105.38	348.31	180	hypothetical protein DKF5p4.9
+	pir2:T47707	7.00	105.27	347.43	183	hypothetical protein F1116.9
+	pir2:S31894	7.00	105.13	346.29	187	probable iron-sulfur prote
+	pir2:B42957	7.00	105.03	345.45	190	iron-sulfur protein Coof [si
+	pir2:F72559	7.00	104.96	344.90	192	hypothetical protein APE1760
+	pir2:A69375	7.00	104.89	344.35	194	hypothetical protein AF1001
+	pir2:E71064	7.00	104.89	344.35	194	hypothetical protein A - Bac
+	pir2:E75297	7.00	104.89	344.35	194	hypothetical protein PH211
+	pir2:TC5613	7.00	104.89	344.35	194	hypothetical protein - Beino
+	pir2:H83297	7.00	104.75	343.28	198	ryudocan precursor - mouse
+		7.00	104.75	343.28	198	hypothetical protein PA2792



pir2:G75020	+	7.00	99.80	306.17	419	glucose-1-phosphate thymidyllyl	pir2:JC4534	-	7.00	98.16	294.79	537	cytochrome P450 4F6 protein
pir2:S42989	+	7.00	99.80	306.17	419	T48 protein - fruit fly (Drosop	pir2:S40111	+	7.00	98.14	294.62	539	capsid protein - human calici
pir2:F83494	+	7.00	99.79	306.06	420	hypothetical protein PA1220 (im	pir2:C70732	-	7.00	98.10	294.37	542	probable integral membrane p
pir2:B70178	+	7.00	99.74	305.73	423	vacuolar X-prolyl dipeptidyl am	pir2:F83202	-	7.00	98.09	294.29	543	alginatase-c5-mannuronan-epim
pir2:E71982	+	7.00	99.71	305.51	425	isocitrate dehydrogenase - Heil	pir1:PWBYA	+	7.00	98.06	294.13	545	H+-transporting ATP synthase
pir2:C64523	+	7.00	99.71	305.51	425	isocitrate dehydrogenase (NADP+	pir2:T02029	+	7.00	98.05	294.04	546	DNA-binding protein pabf - (
pir2:S04848	+	7.00	99.69	305.40	426	transcription initiation factor	pir2:T15523	+	7.00	98.03	293.88	548	hypothetical protein C1688.
pir2:D75582	+	7.00	99.66	305.18	428	hypothetical protein - Deinococ	pir2:B70508	+	7.00	98.02	293.80	549	probable htrA protein - Myc
pir2:E82583	-	7.00	99.63	304.96	430	hypothetical protein - Deinococ	pir2:T35451	+	7.00	97.97	293.47	553	ABC transporter integral mem
pir2:T31539	+	7.00	99.62	304.96	430	threonine synthase XF2223 (impc	pir2:S45126	+	7.00	97.96	293.39	554	hemolysin accessory fhaC pro
pir2:T35519	+	7.00	99.39	303.27	441	probable two-component system	pir2:C83233	+	7.00	97.86	292.75	562	long-chain-fatty-acid-CoA
pir1:FE4637	+	7.00	99.37	303.16	447	hypothetical protein Y47D3A.7	pir1:CZCLBM	+	7.00	97.85	292.67	563	cellulase (EC 3.2.1.4) B pr
pir2:H64741	+	7.00	99.33	302.85	450	hypothetical protein SC6610.100	pir2:E70752	+	7.00	97.85	292.67	563	probable dead protein - Myc
pir2:T47824	+	7.00	99.33	302.85	450	D-alanine/glycine transport pro	pir2:T07850	-	7.00	97.81	292.43	566	leucyl aminopeptidase (EC 3
pir2:G36051	+	7.00	99.32	302.75	451	yael protein - Escherichia coli	pir2:T07850	-	7.00	97.81	292.43	566	proteol disulfide-isomerase
pir2:T21822	+	7.00	99.32	302.75	451	hypothetical protein F24G16.270	pir2:T11653	+	7.00	97.81	292.43	566	hypothetical protein SPAC6Fp
pir2:F82307	+	7.00	99.26	302.34	455	trypsin-like proteinase (EC 3.4	pir2:T11653	+	7.00	97.81	292.43	566	leucyl aminopeptidase (EC 3
pir2:T21822	+	7.00	99.24	302.24	456	proteinase DO VC0566 (imported)	pir2:S57812	+	7.00	97.78	292.20	569	urease (EC 3.5.1.5) alpha ch
pir2:T21822	+	7.00	99.24	302.24	456	hypothetical protein F35H8.5 -	pir2:S75169	+	7.00	97.78	292.20	569	C conserved hypothetical prote
pir2:B39200	+	7.00	99.23	302.14	457	siroheme synthase - Salmonella	pir2:A69511	-	7.00	97.78	292.20	569	leucyl aminopeptidase (EC 3
pir2:B72745	+	7.00	99.21	302.04	458	probable pyruvate kinase APE048	pir2:T07849	-	7.00	97.76	292.04	571	beta-fructofuranosidase (EC
pir2:T36334	+	7.00	99.20	301.94	459	dihydrolipoamide dehydrogenase	pir1:S41376	+	7.00	97.76	292.04	571	conserved hypothetical prote
pir2:T10995	+	7.00	99.18	301.84	460	hypothetical protein SCF9.15c -	pir2:T08439	+	7.00	97.73	291.89	573	leucyl aminopeptidase (EC 3
pir2:T16771	+	7.00	99.18	301.84	460	NADH dehydrogenase (ubiquinone)	pir2:T08439	+	7.00	97.73	291.89	573	leucyl aminopeptidase (EC 3
pir2:T34936	+	7.00	99.18	301.84	460	phosphomannomutase (EC 5.4.2.8)	pir2:T08439	+	7.00	97.73	291.89	573	beta-fructofuranosidase (EC
pir2:S75657	-	7.00	99.11	301.34	465	hypothetical protein SC3F9.14 S	pir2:T31675	+	7.00	97.57	290.81	587	conserved hypothetical prote
pir1:A48788	-	7.00	99.10	301.24	466	hypothetical protein SC3F9.14 S	pir2:T31675	+	7.00	97.57	290.81	587	methy-accepting chemotaxis
pir2:A24450	-	7.00	99.06	300.95	469	coproporphyrinogen oxidase (EC	pir2:T35549	-	7.00	97.57	290.81	587	hypothetical protein SC6G3.C
pir2:T35526	+	7.00	99.06	300.95	469	collagen alpha 2(VIII) chain -	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein - fruit
pir2:E83195	+	7.00	99.06	300.95	469	probable integral membrane tran	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:D70542	-	7.00	99.06	300.95	469	probable MFS transporter PA3595	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:C70853	-	7.00	99.01	300.65	472	hypothetical protein Rv1586c -	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T23793	-	7.00	98.97	300.36	475	hypothetical protein Rv3087 - M	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:A70318	-	7.00	98.96	300.27	476	aldehyde dehydrogenase - Aquif	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T43042	+	7.00	98.92	300.08	478	hypothetical protein - fission	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:A71273	+	7.00	98.92	299.98	479	hypothetical protein - fission	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir1:T34809	+	7.00	98.89	299.79	481	methylophosphate mutase (EC 5.4	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T23654	+	7.00	98.88	299.69	482	probable leucyl aminopeptidase	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:H70112	+	7.00	98.86	299.60	483	ARS-binding factor-1 - yeast (K	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T25684	+	7.00	98.84	299.41	485	periplasmic serine proteinase	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T50749	+	7.00	98.84	299.41	485	hypothetical protein F08D12.12	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S49624	+	7.00	98.82	299.32	486	methoxynurosporene dehydrogena	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:D70876	+	7.00	98.82	299.32	486	methoxynurosporene dehydrogena	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:A82984	-	7.00	98.79	299.13	488	probable polyketidesynthase bel	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir1:I37062	+	7.00	98.70	299.13	488	hypothetical protein PA5294 (im	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S23633	+	7.00	98.70	298.48	495	involucrin S - gorilla	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T46176	+	7.00	98.66	298.20	498	methoxynurosporene dehydrogena	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T46176	+	7.00	98.66	298.20	498	probable cytochrome P450 T4D2.2	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T70012	+	7.00	98.63	298.02	500	leucyl aminopeptidase homolog	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir1:KICRGE	+	7.00	98.57	297.57	505	phosphoglycerate kinase (EC 2.7	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T46996	+	7.00	98.57	297.57	505	hypothetical protein arud (impc	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:A56391	+	7.00	98.55	297.57	506	lamina associated polypeptide 1	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:B70368	-	7.00	98.50	297.48	510	L-aspartate oxidase - Aquifex a	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S10527	+	7.00	98.49	297.03	511	endoglucanase B precursor - Pse	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:B71722	+	7.00	98.46	296.86	513	probable periplasmic serine pro	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T40998	+	7.00	98.46	296.86	513	hypothetical protein SPCC1450.1	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:A82432	+	7.00	98.46	296.86	513	hypothetical protein SPCC1450.1	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:B83360	+	7.00	98.45	296.77	514	sodium/solute symporter VCA0667	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:A82634	+	7.00	98.45	296.77	514	hypothetical protein PA2286 (im	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:H69034	+	7.00	98.39	296.33	519	2-isopropylmalate synthase XF18	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:D70695	-	7.00	98.39	296.33	519	conserved hypothetical protein	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S78811	-	7.00	98.34	295.98	523	leucyl aminopeptidase (EC 3.4.1	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T07834	+	7.00	98.34	295.98	523	hypothetical protein Rv3776 - M	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:JM0902	+	7.00	98.31	295.81	525	hydroxymethylpyrimidine kinase	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T49212	-	7.00	98.29	295.64	527	pituitary adenylate cyclase act	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T37572	+	7.00	98.29	295.64	527	hypothetical protein F27K19.180	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S36520	+	7.00	98.27	295.55	528	hypothetical protein ZC434.9 -	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:D65214	+	7.00	98.27	295.55	528	l1 protein - human papillomavir	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:J00783	-	7.00	98.26	295.47	529	hypothetical 60.8 kD protein in	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S62439	-	7.00	98.25	295.38	530	55.5K sporulation protein - Str	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:T45197	+	7.00	98.21	295.13	533	hypothetical serine rich protein	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir2:S44827	-	7.00	98.19	294.96	535	probable serine proteinase htrA	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
pir1:UTADIA	+	7.00	98.17	294.88	536	F54F2.2 protein - Caenorhabdit	pir2:T35549	-	7.00	97.56	290.74	588	hypothetical protein SC7A1.C
						noncapsid protein NS1 - adeno-a	pir2:S57984	-	7.00	95.11	274.74	852	probable membrane protein YD

```

pir2:S43846      7.00 95.05 274.35      +      xylanase B - rumen fungus (Neoc
pir2:T12570      7.00 94.91 273.67      +      hypothetical protein C23F12.2 -
pir2:T12621      7.00 94.91 273.48      +      hypothetical protein F32AF7.5 -
pir2:T72397      7.00 94.88 273.34      +      pyruvate,orthophosphate dikinas
pir2:PC1231      7.00 94.87 273.29      +      hypothetical protein 882 (3' re
pir2:T38387      7.00 94.86 273.15      +      hypothetical protein SPAC25H1.0
pir2:S54355      7.00 94.85 273.10      +      inter-alpha-trypsin inhibitor 1
pir2:T03081      7.00 94.84 273.01      +      hypothetical protein T10P11.3.2
pir2:T23229      7.00 94.71 272.22      +      hypothetical protein ZC101.1 -
pir2:C70168      7.00 94.69 272.08      +      DNA polymerase I (polA) homolog
pir2:G54347      7.00 94.63 271.72      +      SNF2/Rad54 helicase-related pro
pir2:S54342      7.00 94.61 271.45      +      protein-tyrosine-phosphatase (P
pir2:T37256      7.00 94.59 271.45      +      metalloproteinase sup-17 - Caen
pir2:S28394      7.00 94.47 270.69      +      probable serine/threonine-speci
pir2:S27921      7.00 94.42 270.38      +      nuclear antigen EBNA-3B - human
pir2:T15342      7.00 94.30 269.61      +      hypothetical protein B0336.1 -
pir2:T00353      7.00 94.27 269.44      +      hypothetical protein KIAA0697 -
pir1:JC2349      7.00 94.19 268.97      +      protein-tyrosine-phosphatase (P
pir2:T48721      7.00 94.08 268.27      +      PTP 35 protein - mouse
pir2:T28654      7.00 94.03 267.98      +      transposase - Pseudomonas puti
pir2:PN0156      7.00 94.01 267.82      +      glutamate receptor channel del
pir2:S28858      7.00 94.00 267.78      +      glutamate receptor delta-2 cha
pir2:T42229      7.00 93.92 267.29      +      probable EL-E2 ATPase (EC 3.6.
pir2:S44644      7.00 93.92 267.25      +      hypothetical protein F37A4.7 -
pir2:T42093      7.00 93.83 266.70      +      phospholipase D (EC 3.1.4.4) 1
pir2:T13732      7.00 93.82 266.66      +      phospholipase D (EC 3.1.4.4) 1
pir2:T18530      7.00 93.82 266.66      +      phospholipase D (EC 3.1.4.4) 1
pir2:T34258      7.00 93.82 266.66      +      phospholipase D (EC 3.1.4.4) 1
pir2:T13943      7.00 93.81 266.62      +      phospholipase D (EC 3.1.4.4) 1
pir2:T29092      7.00 93.80 266.50      +      TSC-22 protein homolog - chick
pir1:YGVCAR      7.00 93.74 266.19      +      angr protein - Vibrio anguilla
pir2:T23764      7.00 93.74 266.19      +      hypothetical protein M142.6 -
pir2:A27079      7.00 93.74 266.15      +      fibronectin receptor alpha cha
pir2:S24257      7.00 93.71 266.00      +      integrin alpha-5 chain - mouse
pir2:D81826      7.00 93.61 265.39      +      exodeoxyribonuclease V (EC 3.1
pir2:A81050      7.00 93.61 265.39      +      exodeoxyribonuclease V 125 kd
pir2:T42681      7.00 93.61 265.39      +      hypothetical protein DKFp434B
pir2:T11203      7.00 93.58 265.20      +      phospholipase (EC 3.1.4.4.) D1
pir2:T13725      7.00 93.58 265.20      +      phospholipase D (EC 3.1.4.4) 1
pir2:T46635      7.00 93.58 265.16      +      phospholipase D (EC 3.1.4.4) 1
pir2:T20579      7.00 93.56 265.08      +      hypothetical protein F46C3.1 -
pir2:T14576      7.00 93.49 264.64      +      nosA protein - slime mold (Dic
pir2:T43275      7.00 93.45 264.41      +      neurabin - rat
pir2:C72409      7.00 93.40 264.08      +      reverse gyrase - Thermotoga ma
pir2:A43253      7.00 93.36 263.87      +      large tra-1 protein - Caenorha
pir2:S63399      7.00 93.32 263.61      +      probable membrane protein YNRQ
pir2:S54504      7.00 93.30 263.47      +      hypothetical protein YPR030w -
pir2:B41206      7.00 93.28 263.33      +      microtubule-associated protein
pir2:T49403      7.00 93.24 263.08      +      related to protein-tyrosine-ph
pir2:S54496      7.00 93.23 263.04      +      probable membrane protein YPRQ
pir2:S20106      7.00 93.12 262.37      +      hypothetical protein 2 - slime
pir2:A33183      7.00 93.12 262.37      +      microtubule-associated protein
pir2:I50620      7.00 93.00 261.65      +      prockr2 - chicken (fragment)
pir2:T04584      7.00 92.97 261.45      +      TMV resistance protein N homol
pir2:S47536      7.00 92.90 261.08      +      SWH1 protein - yeast (Saccharo
pir2:A53491      7.00 92.90 261.04      +      bumetanide-sensitive Na-K-Cl c
pir2:S26722      7.00 92.88 260.91      +      DNA-directed RNA polymerase (R
pir2:S72620      7.00 92.82 260.55      +      probable reverse transcriptase
pir2:T00332      7.00 92.78 260.35      +      hypothetical protein KIA0559
pir2:S53714      7.00 92.76 260.19      +      sericin1B - silkworm
pir2:S54570      7.00 92.75 260.12      +      probable membrane protein YMRP
pir2:T34929      7.00 92.66 259.60      +      DNA-binding protein - human cy
pir2:T20739      7.00 92.64 259.51      +      hypothetical protein SCF9.07
pir2:T22904      7.00 92.62 259.38      +      hypothetical protein F1A10.1
pir2:T29041      7.00 92.54 258.87      +      hypothetical protein B0228.4 -
pir2:S37958      7.00 92.47 258.47      +      myosin heavy chain homolog YKL
pir2:T00964      7.00 92.10 256.24      +      hypothetical protein F2022.14
pir1:JH0675      7.00 92.06 256.01      +      restrictin precursor - chicken
pir2:VCBE17      7.00 91.96 255.41      +      major capsid protein - human H
pir2:IS56814      7.00 91.84 254.74      +      microtubule-interacting protei
pir2:T42636      7.00 91.73 254.08      +      protein-tyrosine-phosphatase C
pir2:T31153      7.00 91.71 253.99      +      hypothetical protein 235 - Sp
pir1:B48148      7.00 91.64 253.53      +      protein-tyrosine-phosphatase (

```

```

pir1:A48148      7.00 91.62 253.45      +      protein-tyrosine-phosphatas
pir2:T13709      7.00 91.58 253.21      +      diacylglycerol kinase (EC 2.
nucleoporin - rat
pir2:A44345      7.00 91.52 252.84      +      protein-tyrosine-phosphatas
pir2:D41214      7.00 90.83 250.58      +      protein-tyrosine-phosphatas
pir2:CA1214      7.00 90.83 248.84      +      collagen alpha 6(IV) chain
pir1:GSHU6B      7.00 90.58 247.45      +      adenylyate cyclase (EC 4.6.1
pir2:A33988      7.00 90.58 247.42      +      retinoblastoma binding prot
pir1:T38902      7.00 90.47 246.78      +      microtubule-associated sorl
pir2:S54602      7.00 90.42 246.50      +      hypothetical protein YK1014
pir2:S37827      7.00 90.30 245.85      +      hypothetical protein H_GS54
pir2:T00637      7.00 89.74 242.66      +      hypothetical protein KIAA04
pir2:T00093      7.00 89.30 240.19      +      mycoserolate synthase (EC 2
pir2:S27005      7.00 89.10 239.09      +      voltage-dependent calcium c
pir2:A44467      7.00 89.03 238.73      +      voltage-dependent calcium c
pir2:JH0426      7.00 89.02 238.66      +      voltage-dependent calcium c
pir2:TJ0427      7.00 88.95 238.27      +      calcium channel protein - r
pir2:T13339      7.00 88.93 238.19      +      calcium channel alpha-1 cha
pir2:S05054      7.00 88.79 237.38      +      hypothetical protein C52A11
pir2:T20145      7.00 88.79 237.38      +      calcium channel protein typ
pir2:A52290      7.00 88.47 235.68      +      hypothetical protein T05C3.
pir2:T31733      7.00 88.06 233.45      +      zonadhesin - pig
pir2:T34022      7.00 87.92 232.67      +      hypothetical protein R13F6.
pir2:T16743      7.00 87.03 227.93      +      cell proliferation antigen
pir2:B48666      7.00 86.73 226.38      +      hypothetical protein - Syne
pir2:A48666      7.00 86.25 223.90      +      cell proliferation antigen
pir1:A42996      7.00 85.08 223.02      +      genome polyprotein - mosqui
pir2:S43048      7.00 85.74 221.26      +      polyketide synthase type 1
pir2:E83641      7.00 85.71 221.11      +      probable hemagglutinin PA00
pir2:T23433      7.00 85.46 219.83      +      hypothetical protein K08C7.
pir2:T37316      7.00 85.40 219.54      +      probable laminin alpha chai
pir2:S02392      7.00 84.05 212.79      +      alpha-2-macroglobulin recep
pir1:S25111      7.00 84.05 212.78      +      alpha-2-macroglobulin recep
pir2:T17464      7.00 83.33 209.27      +      rifamycin polyketide syntha
pir2:T44807      7.00 82.95 207.44      +      mycosubtilin synthetase cha
pir2:C83339      7.00 82.64 205.96      +      hypothetical protein PA2462
pir2:T14593      7.00 79.27 189.27      +      syringomycin synthetase - P
seq_name: pir2:B81914
seq_documentation_block:
probable periplasmic serine proteinase (EC 3.4.21.-) NMA0710 [imported] - Neisseria m
C:Species: Neisseria meningitidis
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 19-May-2000
C:Accession: B81914
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.K.; Mo
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: A81775; MUID:20222556
A:Accession: B81914
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-499 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NTD:g7379424; PIDN:CAB83996.1; PID:g7371
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0710
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase
alignment_scores:
Quality: 127.00      Length: 127
Ratio: 1.000      Gaps: 0
Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-388-090-3 x B81914      ..
Align seg 1/1 to: B81914      from: 1 to: 499
583 TTGAACCGGGCGAATGGTTCGCTCCATCGGGCGCCCTTCGGCTTTGA 632

```



195 LeuLysProGlyGlyThrProValAlaAlaIleGlyAlaProPheGlyPheAs 211  
|||||  
633 CAACAGCGTGACCGCGCGCATCGTGTCCGCCAAAGCGAGAAGCCTGCCCA 682  
|||||  
211 pAsnSerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProA 228  
|||||  
683 ACGAAAGCTACACACCCCTTCATCCAAACGAGCGTTGCCATCAATCGGGC 732  
|||||  
228 sngLuSerTyrThrProPheIleGlnThrAspValAlaIleAsnProGly 244  
|||||  
733 AATTCCCGCGCGCCCTGTCAACTTAAAGACACAGGTCGTGGCATCAA 782  
|||||  
245 AsnSerGlyGlyProLeuPheAsnLeuLysGlyGlnValGlyIleAs 261  
|||||  
783 TTGCAATATACAGCCGCGCGGATTTCATGGGCATCTCTTTGCCA 832  
|||||  
261 nSerGlnIleYrSerArgSerGlyGlyPheMetGlyIleSerPheAlaI 278  
|||||  
833 TCCGATTGACGTTGCCATGAATGTGCGCGAACAGCTGAAACACCGGC 882  
|||||  
278 leProIleAspValAlaMetAsnValAlaGluGlnLeuLysAsnThrGly 294  
|||||  
883 AAAGTCCCAACCGGCAACTGGCGCTGATTATTCAGGAAGTATCTTACGS 932  
|||||  
295 LysValGlnArgGlyGlnLeuGlyValIleIleGlnGluValSerTyrGl 311  
|||||  
933 TTTGGCACAGTCGTTGCGTCTGGATAAAGCC 963  
|||||  
311 yLeuAlaGlnSerPheGlyLeuAspLysAla 321  
|||||

seq\_name: pir2:F83550

seq\_documentation\_block:  
serine proteinase MucD precursor PA0766 [imported] - Pseudomonas aeruginosa (strain PA01  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 15-Sep-2000  
R:Accession: F83550  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Br  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,  
.; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho  
A:Reference number: A82950  
A:Accession: F83550  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-474 <STO>  
A:Cross-references: GB:AE004511; GB:AE004091; NID:g9946646; PIDN:AAG04155.1; GSPDB:GN001  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: mucD; PA0766

alignment\_scores:  
Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x F83550 ..

Align seg 1/1 to: F83550 from: 1 to: 474

637 AGCGTGACCGCGCGCATCGTGTCCGCCAAAGCGAGAAGCCTGCCCAACGA 686  
|||||  
184 SerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProAsnGl 200  
|||||  
687 AAGCTAC 693  
|||||  
200 uSerTyr 202  
|||||

seq\_name: pir2:F70048

seq\_documentation\_block:  
serine proteinase Do homolog yvtB - Bacillus subtilis  
C:Species: Bacillus subtilis  
C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 15-Oct-1999  
C:Accession: F70048  
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
Nature 390, 249-256, 1997  
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga  
iesch, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma  
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, M.; Rose, F.; Sadale, Y.; Sato, T.; Scan  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid  
A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil  
A:Reference number: A69580; MUID:98044033  
A:Accession: F70048

A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-224 <KUN>  
A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CAB15289.1; PID:el184  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yvtB

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x F70048 ..

Align seg 1/1 to: F70048 from: 1 to: 224

718 GCCATCAATCCGGCAATTCCGGCGGCCCGCTG 750  
|||||  
58 AlaIleAsnProGlyAsnSerGlyGlyProLeu 68

seq\_name: pir2:S74643

seq\_documentation\_block:

proteinase hhoA (EC 3.4.-.-) - Synechocystis sp. (strain PCC 6803)  
N:Alternate names: protein sll1679

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S74643

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,  
O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas  
DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys  
S.

A:Reference number: S74322; MUID:97061201

A:Accession: S74643

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-394 <KAN>

A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAAL6795.1; PID:g165

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: hhoA

C:Superfamily: proteinase hhoB

C:Keywords: hydrolase; proteinase

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x S74643 ..

Align seg 1/1 to: S74643 from: 1 to: 394

718 GCCATCAATCCGGCAATTCGGCGGCCGCGCTG 750

|||||  
231 AlaIleAsnProGlyAsnSerGlyGlyProLeu 241

seq\_name: pir2:S71284

## seq\_documentation\_block:

probable periplasmic serine proteinase DO (htrA-1) - syphilis spirochete

C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)

C:Date: 24-Jul-1998 #sequence\_revision 24-Jul-1998 #text\_change 24-Nov-1999

C:Accession: B71284

R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn

they, L.; Weidman, J.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD

Science 281, 375-388, 1998

A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.

A:Reference number: A71250; MUID:98332770

A:Accession: B71284

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-398 &lt;COL&gt;

A:Cross-references: GB:AE001248; GB:AE000520; NID:g3323074; PIDN:AAC65740.1; PID:g332308

A:Experimental source: strain Nichols

C:Genetics:

A:Gene: TP0773

C:Superfamily: proteinase hhoB

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x B71284 ..

Align seg 1/1 to: B71284 from: 1 to: 398

718 GCCATCAATCCGGCAATTCGGCGGCCGCGCTG 750

|||||  
223 AlaIleAsnProGlyAsnSerGlyGlyProLeu 233

seq\_name: pir2:S75445

## seq\_documentation\_block:

proteinase hhoB (EC 3.4.21.1) - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein sll1427

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S75445

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

s.

A:Reference number: S74322; MUID:97061201

A:Accession: S75445

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-416 &lt;KAN&gt;

A:Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAAL8006.1; PID:g165309

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

## C:Genetics:

A:Gene: hhoB

C:Superfamily: proteinase hhoB

C:Keywords: hydrolase; proteinase

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x S75445 ..

Align seg 1/1 to: S75445 from: 1 to: 416

718 GCCATCAATCCGGCAATTCGGCGGCCGCGCTG 750

|||||  
252 AlaIleAsnProGlyAsnSerGlyGlyProLeu 262

seq\_name: pir2:E75357

## seq\_documentation\_block:

probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000

C:Accession: E75357

R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.

S.; Smith, H.O.; Venter, J.C.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896

A:Accession: E75357

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-441 &lt;WHI&gt;

A:Cross-references: GB:AE002017; GB:AE000513; NID:g6459527; PIDN:AAF11312.1; PID:g645

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR1756

A:Map position: 1

C:Superfamily: proteinase hhoB

## alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x E75357 ..

Align seg 1/1 to: E75357 from: 1 to: 441

718 GCCATCAATCCGGCAATTCGGCGGCCGCGCTG 750

|||||  
256 AlaIleAsnProGlyAsnSerGlyGlyProLeu 266

seq\_name: pir2:A69643

## seq\_documentation\_block:

serine proteinase Do, heat-shock protein htrA - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1987 #sequence\_revision 05-Dec-1987 #text\_change 20-Jun-2000

C:Accession: A69643

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Iulio, P.

Koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: A69643  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-449 <KUN>  
 A:Cross-references: GB:299110; GB:AL009126; NID:g2633472; PIDN:CABI3147.1; PID:g2633644  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: htrA  
 C:Superfamily: proteinase hhoB

alignment\_scores:  
 Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x A69643 ..

Align seg 1/1 to: A69643 from: 1 to: 449

718 GCCATCAATCCGGCAATTCCGGCGCCGCTG 750  
 |||||

284 AlaIleAsnProGlyAsnSerGlyProLeu 294

seq\_name: pir2:S77538

seq\_documentation\_block:

serine proteinase (EC 3.4.21.-) htrA - *Synechocystis* sp. (strain PCC 6803)

N:Alternate names: protein sir1204

C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: S77538

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

S.

A:Reference number: S74322; MUID:97061201

A:Accession: S77538

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-452 <KAN>

A:Cross-references: EMBL:D90905; GB:AB001339; NID:g1652360; PIDN:BAAL7385.1; PID:g165246

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: htrA

C:Superfamily: proteinase hhoB

C:Keywords: hydrolase; serine proteinase

alignment\_scores:  
 Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x S77538 ..

Align seg 1/1 to: S77538 from: 1 to: 452

718 GCCATCAATCCGGCAATTCCGGCGCCGCTG 750  
 |||||

290 AlaIleAsnProGlyAsnSerGlyProLeu 300

seq\_name: pir1:B70426

seq\_documentation\_block:

periplasmic serine proteinase (EC 3.4.21.-) - *Aquifex aeolicus*

C:Species: *Aquifex aeolicus*

C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 21-Jul-2000

C:Accession: B70426

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.,

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A:Reference number: A70300; MUID:98196666

A:Accession: B70426

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-453 <AQF>

A:Cross-references: GB:AE000741; GB:AE000657; NID:g2983841; PIDN:AA07399.1; PID:g298

A:Experimental source: strain VF5

C:Genetics:

A:Gene: htrA

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

alignment\_scores:

Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x B70426 ..

Align seg 1/1 to: B70426 from: 1 to: 453

718 GCCATCAATCCGGCAATTCCGGCGCCGCTG 750

|||||

199 AlaIleAsnProGlyAsnSerGlyProLeu 209

seq\_name: pir2:F72359

seq\_documentation\_block:

periplasmic serine proteinase Do (EC 3.4.21.-) - *Thermotoga maritima* (strain MSB8)

N:Alternate names: heat shock protein htrA

C:Species: *Thermotoga maritima*

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 21-Jul-2000

C:Accession: F72359

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: F72359

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-459 <ARN>

A:Cross-references: GB:AE001732; GB:AE000512; NID:g4981087; PIDN:AA035656.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0571

C:Superfamily: Helicobacter serine proteinase

C:Keywords: hydrolase; serine proteinase

alignment\_scores:

Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x F72359 ..

Align seg 1/1 to: F72359 from: 1 to: 459

718 GCCATCAATCCGGCGCAATTCCGGCGGCCGCTG 750  
|||||  
200 AlalleAsnProGlyAsnSerGlyGlyProLeu 210

seq\_name: plr2:G81528

seq\_documentation\_block:  
serine proteinase, HtrA/DegQ/DegS family CP0877 [imported] - Chlamydothiia pneumoniae (s  
C:Species: Chlamydothiia pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 21-Jul-2000  
C:Accession: G81528  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: G81528  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <REA>  
A:Cross-references: GB:AE002246; GB:AE002161; NID:g7189785; PIDN:AAF38665.1; PID:g718978  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
C:Superfamily: Helicobacter serine proteinase

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x G81528 ..

Align seg 1/1 to: G81528 from: 1 to: 488

718 GCCATCAATCCGGCGCAATTCCGGCGGCCGCTG 750  
|||||  
232 AlalleAsnProGlyAsnSerGlyGlyProLeu 242

seq\_name: plr2:G72011

seq\_documentation\_block:  
do serine proteinase - Chlamydothiia pneumoniae (strain CWL029)  
C:Species: Chlamydothiia pneumoniae, Chlamydia pneumoniae  
C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000  
C:Accession: G72011  
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;  
Nature Genet. 21, 385-389, 1999  
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.  
A:Reference number: A72000; MUID:99206606  
A:Accession: G72011  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <ARN>  
A:Cross-references: GB:AE001678; GB:AE001363; NID:g4377301; PIDN:AAD19116.1; PID:g437730  
A:Experimental source: strain CWL029  
C:Genetics:  
A:Gene: htrA  
C:Superfamily: Helicobacter serine proteinase

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x G72011 ..

Align seg 1/1 to: G72011 from: 1 to: 488

718 GCCATCAATCCGGCGCAATTCCGGCGGCCGCTG 750  
|||||  
232 AlalleAsnProGlyAsnSerGlyGlyProLeu 242

seq\_name: plr2:B81728

seq\_documentation\_block:  
serine proteinase, HtrA/DegQ/DegS family TC0210 [imported] - Chlamydia muridarum (str  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 21-Jul-2000  
C:Accession: B81728  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heideberg, J.F.; White, O.; Hickey,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.  
A:Reference number: A81500; MUID:20150255  
A:Accession: B81728  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <TET>  
A:Cross-references: GB:AE002288; GB:AE002160; NID:g7190247; PIDN:AAF39082.1; PID:g719  
A:Experimental source: strain Nlgg (MoPn)  
C:Genetics:  
C:Superfamily: Helicobacter serine proteinase

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x B81728 ..

Align seg 1/1 to: B81728 from: 1 to: 497

718 GCCATCAATCCGGCGCAATTCCGGCGGCCGCTG 750  
|||||  
241 AlalleAsnProGlyAsnSerGlyGlyProLeu 251

seq\_name: plr2:H71465

seq\_documentation\_block:  
probable do serine proteinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: H71465  
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: H71465  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-497 <ARN>  
A:Cross-references: GB:AE001355; GB:AE001273; NID:g3329292; PIDN:ANC68420.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: htrA  
C:Superfamily: Helicobacter serine proteinase

alignment\_scores:  
Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:  
US-09-388-090-3 x H71465 ..

Align seg 1/1 to: H71465 from: 1 to: 497

718 GCCATCAATCGGGCAATTCCGGGGGCCCGCTG 750  
||||||||||||||||||||||||||||||||  
241 AlaIleAsnProGlyAsnSerGlyGlyProLeu 251



OM of: US-09-388-090-3 to: SwissProt\_39:\* out\_format : pfs

Date: Mar 22, 2001 12:09 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame-n2p.model -DEV=xlpl  
-Q/cgn2\_1/USPT0001/JUS09388090/runat\_20032001-092951\_25871/app\_query.fasta\_1.1474  
-DB=SwissProt\_39 -QPMF=fastan -SUFFIX=oligo7.rsp -GAPOP=4.500  
-CAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -CAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000  
-XGAPOP=6.000 -XGAPEXT=7.000 -XGAPOP=60.000 -XGAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human40.cdi -LIST=1000 -DOALIGN=200 -THR\_SCORE=quality  
-THR\_MIN=7 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=2000000000 -USER=US09388090\_@CGN1\_1\_34 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-388-090-3

Query length: 1395

Database: SwissProt\_39:\*

Database sequences: 88757

Database length: 3294092

Search time (sec): 26.910000

WARN: XGAPOP and XGAPOP must be equal. Assuming XGAPOP=XGAPOP=60.000  
WARN: XGAPEXT and XGAPEXT must be equal. Assuming XGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd Orig	ZScore	Score Len	Documentation
SwissProt_39:DEGP_ARATH	11.00	172.19	0.0130	437
SwissProt_39:DEGP_CHLPP	11.00	171.48	0.0127	438
SwissProt_39:DEGP_CHLPP	11.00	171.36	0.0127	437
SwissProt_39:YXHA_BACSU	9.00	136.16	1.44	400
SwissProt_39:YXHA_BACSU	8.00	120.98	16.34	247
SwissProt_39:YXHA_BACSU	8.00	120.62	16.19	261
SwissProt_39:YXHA_BACSU	8.00	119.72	15.81	300
SwissProt_39:YXHA_BACSU	8.00	119.72	15.81	300
SwissProt_39:YXHA_BACSU	8.00	118.43	15.28	366
SwissProt_39:YXHA_BACSU	8.00	117.41	14.87	429
SwissProt_39:YXHA_BACSU	8.00	116.91	14.68	463
SwissProt_39:YXHA_BACSU	8.00	116.56	14.54	489
SwissProt_39:YXHA_BACSU	8.00	116.38	14.47	503
SwissProt_39:YXHA_BACSU	8.00	115.76	14.24	553
SwissProt_39:YXHA_BACSU	8.00	115.76	14.24	553
SwissProt_39:YXHA_BACSU	8.00	114.92	13.93	630
SwissProt_39:YXHA_BACSU	8.00	114.49	13.77	683
SwissProt_39:YXHA_BACSU	8.00	114.40	13.74	683
SwissProt_39:YXHA_BACSU	8.00	113.15	13.29	829
SwissProt_39:YXHA_BACSU	8.00	110.10	12.27	1328
SwissProt_39:YXHA_BACSU	7.00	113.69	228.43	45
SwissProt_39:YXHA_BACSU	7.00	107.85	195.84	111
SwissProt_39:YXHA_BACSU	7.00	107.79	195.54	112
SwissProt_39:YXHA_BACSU	7.00	106.88	190.88	129
SwissProt_39:YXHA_BACSU	7.00	106.12	187.11	145
SwissProt_39:YXHA_BACSU	7.00	105.99	186.46	148
SwissProt_39:YXHA_BACSU	7.00	105.77	185.41	153
SwissProt_39:YXHA_BACSU	7.00	105.61	184.60	157
SwissProt_39:YXHA_BACSU	7.00	105.09	182.11	170
SwissProt_39:YXHA_BACSU	7.00	104.94	181.39	174
SwissProt_39:YXHA_BACSU	7.00	104.90	181.21	175
SwissProt_39:YXHA_BACSU	7.00	104.37	178.69	190
SwissProt_39:YXHA_BACSU	7.00	104.11	177.44	198
SwissProt_39:YXHA_BACSU	7.00	104.07	177.28	199
SwissProt_39:YXHA_BACSU	7.00	103.98	176.83	202
SwissProt_39:YXHA_BACSU	7.00	103.98	176.83	202
SwissProt_39:YXHA_BACSU	7.00	103.66	175.38	212
SwissProt_39:YXHA_BACSU	7.00	103.34	173.87	223
SwissProt_39:YXHA_BACSU	7.00	103.31	173.74	224

SwissProt_39:NRFC_HAEIN	+	7.00	103.28	173.61	225
SwissProt_39:VATE_DROME	-	7.00	103.25	173.48	226
SwissProt_39:HL_WHEAT	+	7.00	102.97	172.20	236
SwissProt_39:DHSS_PORPU	+	7.00	102.89	171.83	239
SwissProt_39:YGAZ_ECOLI	+	7.00	102.73	171.11	245
SwissProt_39:OCCP_AGRTE	+	7.00	102.29	169.16	262
SwissProt_39:EREL_ARATH	+	7.00	102.20	168.72	266
SwissProt_39:Y893_AQUAE	+	7.00	102.17	168.62	267
SwissProt_39:TRPA_MYCTI	-	7.00	102.08	168.17	271
SwissProt_39:LEPA_MURPS	+	7.00	101.98	167.77	275
SwissProt_39:YARA_PROST	+	7.00	101.82	167.05	282
SwissProt_39:YGS3_YEAST	+	7.00	101.68	166.45	288
SwissProt_39:PFUC_ECOLI	+	7.00	101.59	166.06	292
SwissProt_39:MOXJ_METEX	+	7.00	101.42	165.30	300
SwissProt_39:DDL_BUCAP	-	7.00	101.29	164.74	306
SwissProt_39:Y205_AQUAE	-	7.00	101.23	164.47	309
SwissProt_39:MOCB_RHIME	+	7.00	101.14	164.11	313
SwissProt_39:YHCL_BACSU	-	7.00	101.12	164.02	314
SwissProt_39:PPAC_DEIRA	-	7.00	100.90	163.06	325
SwissProt_39:NECD_MOUSE	+	7.00	100.67	162.05	337
SwissProt_39:ADH1_BACST	+	7.00	100.67	162.05	337
SwissProt_39:G3PX_HORVU	+	7.00	100.63	161.89	339
SwissProt_39:ADH3_BACST	+	7.00	100.63	161.89	339
SwissProt_39:YI21_SYNY3	+	7.00	100.13	159.79	366
SwissProt_39:ALF_PLAFA	-	7.00	100.08	159.37	369
SwissProt_39:P2AI_YEAST	-	7.00	100.08	159.37	369
SwissProt_39:TMAP_AVISA	-	7.00	100.08	159.37	369
SwissProt_39:T186_ECOLI	+	7.00	100.06	159.49	370
SwissProt_39:T421_ECOLI	+	7.00	100.04	159.42	371
SwissProt_39:Y028_ARCFU	+	7.00	100.04	159.42	371
SwissProt_39:P2A2_YEAST	+	7.00	99.94	158.98	377
SwissProt_39:ENTC_ECOLI	+	7.00	99.71	158.00	391
SwissProt_39:ODBB_BOVIN	+	7.00	99.69	157.93	392
SwissProt_39:DOM3_CAEEL	+	7.00	99.67	157.86	393
SwissProt_39:PHV_PSEAE	+	7.00	99.66	157.79	394
SwissProt_39:PHV_PSEFL	+	7.00	99.66	157.79	394
SwissProt_39:PAR2_HUMAN	+	7.00	99.61	157.59	397
SwissProt_39:PAR2_RAT	+	7.00	99.61	157.59	397
SwissProt_39:PAR2_MOUSE	+	7.00	99.57	157.45	399
SwissProt_39:COAT_BBV	+	7.00	99.45	156.92	407
SwissProt_39:YPEB_SNP2	+	7.00	99.38	156.66	411
SwissProt_39:VGLD_HSVBP	+	7.00	99.29	156.27	417
SwissProt_39:VGLD_HSVBP	+	7.00	99.29	156.27	417
SwissProt_39:VGLD_HSVBP	+	7.00	99.27	156.21	418
SwissProt_39:IDH_HELPY	+	7.00	99.17	155.77	425
SwissProt_39:RP54_RHOCA	+	7.00	99.15	155.71	426
SwissProt_39:YAEI_ECOLI	+	7.00	98.80	154.26	450
SwissProt_39:DEQ_ECOLI	+	7.00	98.72	153.97	455
SwissProt_39:RFK9_ECOLI	+	7.00	98.71	153.91	456
SwissProt_39:CYSG_SALTY	+	7.00	98.70	153.85	457
SwissProt_39:HEMN_SYNY3	-	7.00	98.57	153.34	466
SwissProt_39:YU87_MYCTU	-	7.00	98.49	153.25	486
SwissProt_39:BAFI_KUULA	+	7.00	98.30	152.01	472
SwissProt_39:CRTD_RHOSH	+	7.00	98.18	151.77	495
SwissProt_39:AMPA_BACSU	-	7.00	98.11	151.51	500
SwissProt_39:PGKA_RCIFA	-	7.00	98.05	151.26	505
SwissProt_39:GUNE_PSEFL	+	7.00	97.97	150.95	511
SwissProt_39:GUNE_RICPR	+	7.00	97.95	150.85	513
SwissProt_39:VLLI_HPV34	+	7.00	97.76	150.11	528
SwissProt_39:YJQC_ECOLI	+	7.00	97.76	150.11	528
SwissProt_39:SPI5_STRGR	-	7.00	97.75	150.06	529
SwissProt_39:YAY9_SCHPO	+	7.00	97.74	150.01	530
SwissProt_39:ML51_HUMAN	+	7.00	97.69	149.82	534
SwissProt_39:YMA2_CAEEL	-	7.00	97.68	149.77	535
SwissProt_39:VNCA_RAV2	+	7.00	97.67	149.73	536
SwissProt_39:CPPE_RAT	-	7.00	97.65	149.68	537
SwissProt_39:COAT_LORDV	+	7.00	97.63	149.58	539
SwissProt_39:YJ87_MYCTU	+	7.00	97.59	149.44	542
SwissProt_39:ATPA_YEAST	+	7.00	97.56	149.30	545
SwissProt_39:PUT2_AGABI	+	7.00	97.55	149.26	546
SwissProt_39:DEAD_MYCTU	+	7.00	97.35	148.48	563
SwissProt_39:GUNE_CLOTES	+	7.00	97.35	148.48	563
SwissProt_39:AMF2_LYCES	-	7.00	97.28	148.21	569

SwissProt_39:UREL_SYN3 +	7.00	97.28	148.21	569	P73061	synecocystis sp. (su
SwissProt_39:AMPL_LYCES -	7.00	97.26	148.12	571	I10712	lycopersicon esculent
SwissProt_39:AMPL_SOLITU -	7.00	97.23	148.03	573	P13427	solanum tuberosum (pc
SwissProt_39:FIAC_BORPE +	7.00	97.11	147.55	584	P34327	bordeletia pertussis
SwissProt_39:YKTA_CAEEL +	7.00	97.04	147.30	590	I343022	caenorhabditis elegans
SwissProt_39:EVAL_MOUSE +	7.00	97.03	147.25	591	I97767	mus musculus (mouse)
SwissProt_39:YG2S_YEAST -	7.00	96.88	146.67	605	P53261	saccharomyces cerevis
SwissProt_39:YG3F_YEAST -	7.00	96.79	146.30	614	P53283	saccharomyces cerevis
SwissProt_39:ILVD_STRO +	7.00	96.75	146.18	617	I06198	streptomyces coelicol
SwissProt_39:CA28_HUMAN +	7.00	96.57	145.46	635	I25067	homo sapiens (human)
SwissProt_39:DNAK_FRATU +	7.00	96.50	145.19	642	I48205	francisella tularensi
SwissProt_39:YK1L_YEAST +	7.00	96.45	145.00	647	I40489	saccharomyces cerevis
SwissProt_39:YDEE_SCHPO +	7.00	96.44	144.96	648	I10447	schizosaccharomyces p
SwissProt_39:M110_CAEEL +	7.00	96.42	144.88	650	I34400	caenorhabditis elegans
SwissProt_39:GNP1_YEAST -	7.00	96.29	144.40	663	I48813	saccharomyces cerevis
SwissProt_39:FLAY_CAUCR +	7.00	96.14	143.81	679	I15346	caulobacter crescent
SwissProt_39:NEUL_RABIT -	7.00	95.90	142.93	704	I42675	oryzctolagus cuniculu
SwissProt_39:YCCS_ECOLI +	7.00	95.76	142.38	720	I75870	escherichia coli. hy
SwissProt_39:SUF_DROME +	7.00	95.64	141.95	733	I25991	drosophila melanogast
SwissProt_39:FCUA_YEREN +	7.00	95.42	141.14	758	I05202	yersinia enterocoliti
SwissProt_39:MUTS_THECA -	7.00	94.94	139.34	817	I0921x6	thermus aquaticus (su
SwissProt_39:ATCU_RHLIV +	7.00	94.75	138.66	841	I095v3	rhizobium leguminosa
SwissProt_39:VGLH_HSVBC +	7.00	94.74	138.63	842	I27599	bovine herpesvirus ty
SwissProt_39:DEXT_STRUU +	7.00	94.68	138.41	850	I05443	streptococcus mutans
SwissProt_39:SSV1_YEAST -	7.00	94.67	138.35	852	I03770	saccharomyces cerevis
SwissProt_39:OSHI_YEAST -	7.00	94.61	138.16	859	I35845	saccharomyces cerevis
SwissProt_39:FRPP_RAT +	7.00	94.47	137.62	879	I062786	rattus norvegicus (rat)
SwissProt_39:ITH3_MOUSE +	7.00	94.41	137.43	886	I051408	mus musculus (mouse)
SwissProt_39:DPOL_BORBU +	7.00	94.26	136.86	908	I051498	borrelia burgdorferi
SwissProt_39:ST20_YEAST -	7.00	94.04	136.08	939	I03497	saccharomyces cerevis
SwissProt_39:MYG_CANAL +	7.00	93.99	135.90	946	I074254	candida albicans (yea
SwissProt_39:KF5C_MOUSE +	7.00	93.92	135.66	956	I28738	mus musculus (mouse)
SwissProt_39:PTPN_MOUSE +	7.00	93.77	135.11	979	I060673	mus musculus (mouse)
SwissProt_39:PTPN_RAT +	7.00	93.74	135.02	983	I063259	rattus norvegicus (rat)
SwissProt_39:YPT7_CAEEL +	7.00	93.50	134.15	1021	I41885	caenorhabditis eleg
SwissProt_39:ANGR_VIBAN -	7.00	93.33	133.55	1048	I19828	vibrio anguillarum.
SwissProt_39:ITAS_HUMAN +	7.00	93.32	133.53	1049	I08648	homo sapiens (human)
SwissProt_39:ITAS_XENIA +	7.00	93.32	133.51	1050	I06274	xenopus laevis (afri
SwissProt_39:ITAS_MOUSE +	7.00	93.30	133.44	1053	I11688	mus musculus (mouse)
SwissProt_39:PLD1_HUMAN -	7.00	93.17	133.00	1074	I01393	homo sapiens (human)
SwissProt_39:PLD1_RAT -	7.00	93.16	132.97	1075	I70496	rattus norvegicus (rat)
SwissProt_39:TRAI_CAEEL +	7.00	92.96	132.25	1110	I34708	caenorhabditis eleg
SwissProt_39:YN96_YEAST +	7.00	92.92	132.11	1117	I33753	saccharomyces cerevi
SwissProt_39:MAP4_MOUSE +	7.00	92.87	131.95	1125	I27546	mus musculus (mouse)
SwissProt_39:MAP4_HUMAN +	7.00	92.72	131.41	1152	I27816	homo sapiens (human)
SwissProt_39:NIFJ_ENTAG -	7.00	92.60	131.01	1173	I19543	enterobacter agglome
SwissProt_39:NKCL_SQUAC -	7.00	92.50	130.67	1191	I05013	squalus acanthias (s
SwissProt_39:RPOB_THEAC +	7.00	92.48	130.60	1195	I003587	thermoplasma acidoph
SwissProt_39:YMW9_YEAST -	7.00	92.35	130.15	1219	I00439	saccharomyces cerevi
SwissProt_39:DNBI_ICMVA +	7.00	92.27	129.87	1235	I17147	human cytomegaloviru
SwissProt_39:TOP2_LEICH -	7.00	92.24	129.78	1240	I061078	leishmania chagasi.
SwissProt_39:TBP7_CAEEL +	7.00	92.23	129.74	1242	I54816	caenorhabditis eleg
SwissProt_39:MY53_YEAST -	7.00	92.07	129.20	1273	I36006	saccharomyces cerevi
SwissProt_39:VCAP_HSV11 -	7.00	91.58	127.52	1374	I06491	herpes simplex virus
SwissProt_39:MHPI_YEAST -	7.00	91.46	127.15	1398	I43638	saccharomyces cerevi
SwissProt_39:PTPG_CHICK +	7.00	91.35	126.78	1422	I089936	gallus gallus (chick
SwissProt_39:PTPG_MOUSE +	7.00	91.26	126.48	1442	I05909	mus musculus (mouse)
SwissProt_39:PTPG_HUMAN +	7.00	91.25	126.43	1445	I23470	homo sapiens (human)
SwissProt_39:KOG6_DROME +	7.00	91.21	126.30	1454	I09103	drosophila melanog
SwissProt_39:N153_RAT +	7.00	91.15	126.09	1468	I49791	rattus norvegicus (rat
SwissProt_39:PTP1_DROME +	7.00	90.47	123.86	1630	I35992	drosophila melanog
SwissProt_39:CA64_HUMAN -	7.00	90.28	123.25	1678	I14031	homo sapiens (human)
SwissProt_39:CYAA_SCHPO -	7.00	90.23	123.08	1692	I14605	schizosaccharomyces
SwissProt_39:YK84_YEAST +	7.00	89.96	122.21	1764	I34241	saccharomyces cerevi
SwissProt_39:CCAC_MOUSE +	7.00	89.71	118.26	2139	I01815	mus musculus (mouse)
SwissProt_39:CCAC_RAT +	7.00	88.62	117.97	2169	I22002	rattus norvegicus (rat
SwissProt_39:CCAC_RABIT +	7.00	88.62	117.96	2171	I15381	oryzctolagus cuniculu
SwissProt_39:STCA_EMENI +	7.00	88.59	117.86	2181	I12397	emericella nidulans
SwissProt_39:CCAC_HUMAN +	7.00	88.47	117.50	2221	I013936	homo sapiens (human)
SwissProt_39:ZAN_PIG -	7.00	87.77	115.34	2476	I028983	sus scrofa (pig). zonad
SwissProt_39:CCAA_HUMAN +	7.00	87.69	115.11	2505	I000555	homo sapiens (human)
SwissProt_39:K167_HUMAN +	7.00	86.00	110.08	3256	I46013	homo sapiens (human)
SwissProt_39:POLG_MCFA +	7.00	85.83	109.60	3341	I33515	m genome polyprotein

SwissProt\_39:OL56\_STRAT - 7.00 85.49 108.63 3519 | 007017 streptomyces anti  
SwissProt\_39:LM2\_CAEEL - 7.00 85.22 107.85 3672 | 021313 caenorhabditis el  
SwissProt\_39:LRP1\_HUMAN - 7.00 83.84 104.00 4544 | 007954 homo sapiens (hum

seq\_name: SwissProt\_39:DEGP\_ARATH

seq\_documentation\_block:  
ID DEGP\_ARATH STANDARD: PRT: 437 AA.

OC 022609;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROTEASE DO-LIKE PRECURSOR (EC 3.4.21.-).  
GN DEGP.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
OC Magnoliophyta; eudicotyledons; core eudicots; rosidae; eurosids II;  
OC Brassicaceae; Brassicaceae; Arabidopsis.  
[1]  
RN SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RP MEDLINE=98175982; PubMed=9507020;  
RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;  
RT "Identification and characterization of DegP, a serine protease  
RT associated with the luminal side of the thylakoid membrane.";  
RL J. Biol. Chem. 273:7094-7098(1998).  
[2]  
RN SEQUENCE OF 104-118.  
RP Kieselbach T., Bystedt M., Schroeder W.P.;  
RA Submitted (JUL-2000) to the SWISS-PROT data bank.  
CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.  
CC MAY BE INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IN VIVO,  
CC CAN DEGRADE BETA-CASEIN.  
CC -1- ENZYME REGULATION: INHIBITED BY PHENYL METHYL SULFONYL FLUORIDE AND  
CC O-PHENANTHROLINE.  
CC -1- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID  
CC MEMBRANE.  
CC -1- INDUCTION: HEAT SHOCK.  
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
CC DEGP/DEGO/DEGS FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch))  
CC -----  
DR EMBL: AF028842; AAC39436.1; -  
DR MEROPS: S01.279; -  
DR MENDEL: 25267; Arath; 3069; 25267.  
DR INTERPRO: IPR001254; -  
DR INTERPRO: IPR001478; -  
DR INTERPRO: IPR001940; -  
DR PFAM: PF00089; trypsin; 1.  
DR PFAM: PF00595; PD2; 1.  
DR PRINTS: PR00834; PROTEASES2C.  
KW Hydrolyase; Serine protease; Transist peptide; Chloroplast.  
FT TRANSIT 1 103  
FT CHAIN 104 437  
FT ACT\_SITE 171 171  
FT ACT\_SITE 201 201  
FT ACT\_SITE 280 280  
SQ SEQUENCE 437 AA; 46213 MW; 1497B1AB3F5FF2A4 CRC64;

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x DEGP\_ARATH



Align seg 1/1 to: DEGP\_ARATH from: 1 to: 437

718 GCCATCAATCGGGCAATTCCGGCGCGCGCTG 750  
|||||  
274 AlalIeAsnProGlyAsnSerGlyGlyProLeu 284

seq\_name: SwissProt\_39:DEGP\_CHLNP

seq\_documentation\_block:

ID DEGP\_CHLNP STANDARD; PRT; 488 AA.  
AC Q926T0;  
DT 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROBABLE SERINE PROTEASE DO-LIKE PRECURSOR (EC 3.4.21.-).  
GN DEGP OR HTRA OR CPN0979.  
OS Chlamydia pneumoniae (Chlamydia pneumoniae).  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-CWL029;  
RX MEDLINE=99206606; PubMed=10192388;  
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,  
Olinger L., Grimwood J., Davis R.W., Stephens R.S.;  
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis";  
RL Nat. Genet. 21:385-389(1999).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
DEGP/DEGO/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE001678; AAD19116.1; -;  
CC MEROPS; S01.273; -;  
CC INTERPRO; IPR001254; -;  
CC INTERPRO; IPR001478; -;  
CC INTERPRO; IPR001940; -;  
CC PFAM; PF00089; trypsin; 1.  
CC PFAM; PF00595; PDZ; 2.  
CC PRINTS; PR00834; PROTEASES2C.  
CC Hydrolyase; Serine protease; Signal.  
CC SIGNAL 1 20  
CC FT CHAIN 1 20  
CC FT DOMAIN 119 280  
CC FT DOMAIN 281 372  
CC FT DOMAIN 388 476  
CC FT ACT\_SITE 134 134  
CC FT ACT\_SITE 164 164  
CC FT ACT\_SITE 238 238  
CC SEQUENCE 488 AA; 52311 MW; 0BE7E0F88F106F49 CRC64;

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x DEGP\_CHLNP ..

Align seg 1/1 to: DEGP\_CHLNP from: 1 to: 488

718 GCCATCAATCGGGCAATTCCGGCGCGCGCTG 750  
|||||  
232 AlalIeAsnProGlyAsnSerGlyGlyProLeu 242

seq\_name: SwissProt\_39:DEGP\_CHLTR

seq\_documentation\_block:

ID DEGP\_CHLTR STANDARD; PRT; 497 AA.  
AC P18584; O84830;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PROBABLE SERINE PROTEASE DO-LIKE PRECURSOR (EC 3.4.21.-) (59 KDA  
IMMUNOGENIC PROTEIN) (SK59).  
GN DEGP OR HTRA OR CT823.  
OS Chlamydia trachomatis.  
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SEROVAR L2;  
RX MEDLINE=90337348; PubMed=2379836;  
RA Kahane S., Weinstein Y., Sarov I.;  
RT "Cloning, characterization and sequence of a novel 59-kDa protein of  
Chlamydia trachomatis";  
RL Gene 90:61-67(1990).  
CC [2]  
CC SEQUENCE FROM N.A.  
CC STRAIN-D/UW-3/CX;  
CC MEDLINE=99000809; PubMed=9784136;  
CC Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,  
CC Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,  
CC Davis R.W.;  
CC "Genome sequence of an obligate intracellular pathogen of humans:  
Chlamydia trachomatis";  
CC Science 282:754-759(1998).  
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
DEGP/DEGO/DEGS FAMILY.  
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC -!- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS  
CC ON THE AUTHORS TRANSLATED THEIR PUTATIVE 59 KDA IMMUNOGENIC PROTEIN  
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL  
CC PROTEIN.  
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AE001355; AAC68420.1; -;  
CC EMBL; M31119; AAA23116.1; -;  
CC INTERPRO; IPR00126; -;  
CC INTERPRO; IPR001254; -;  
CC INTERPRO; IPR001478; -;  
CC INTERPRO; IPR001940; -;  
CC PFAM; PF00595; PDZ; 2.  
CC PFAM; PF00089; trypsin; 1.  
CC PRINTS; PR00834; PROTEASES2C.  
CC PRINTS; PR00839; V8PROTEASE.  
CC Hydrolyase; Serine protease; Signal; Antigen.  
CC SIGNAL 1 16  
CC FT CHAIN 17 497  
CC FT DOMAIN 128 289  
CC FT DOMAIN 290 381  
CC FT DOMAIN 394 485  
CC FT ACT\_SITE 143 143  
CC FT ACT\_SITE 173 173  
CC FT ACT\_SITE 247 247  
CC SEQUENCE 497 AA; 53244 MW; 86A5E31BB84A38BA CRC64;

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

```
alignment_block:
US-09-388-090-3 x DEGP_CHLTR
Align seg 1/1 to: DEGP_CHLTR from: 1 to: 497

718 GCCATCAATCCGGCAATTCGGCGGCCGCTG 750
|||||
241 AlaileAsnProGlyAsnSerGlyGlyProLeu 251

seq_name: SwissProt_39:YYXA_BACSU

seq_documentation_block:
ID YYXA_BACSU STANDARD; PRT; 400 AA.
AC P39668;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEASE IN ROCR-PURA INTERGENIC REGION (BC 3.4.21.-).
GN YYXA OR YICK.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:168;
RA Kasahara Y., Nakai S., Yoshikawa H., Ogasawara N.;
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 317-400 FROM N.A.
RC STRAIN:168 / MARRBURG;
RX MEDLINE-94156824; PubMed-8113162;
RA Calogero S., Gardan R., Glaser P., Schweitzer J., Rapoport G.,
RA Debarbouille M.;
RT "Rocr", a novel regulatory protein controlling arginine utilization in
RT Bacillus subtilis, belongs to the Ntrc/Nrfa family of transcriptional
RT activators.";
RL J. Bacteriol. 176:1234-1241(1994).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D78193; BA011295.1; -
CC DR EMBL; L22006; -; NOT ANNOTATED_CDS.
CC DR EMBL; Z99124; CAB16073.1; -
CC DR MEROPS; S01.273; -
CC DR SUBTILIST; BG11054; YYXA.
CC DR INTERPRO; IPR001254; -
CC DR INTERPRO; IPR001478; -
CC DR INTERPRO; IPR001940; -
CC DR PFAM; PF00595; PDZ; 1.
CC DR PFAM; PF00089; trypsin; 1.
CC DR PRINTS; PR00834; PROTEASES2C.
KW Hypothetical protein; Hydrolase; Serine protease.
FT ACT_SITE 129 129 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 159 159 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 243 243 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 400 AA; 42788 MW; 91A6E0E897B6F64C CRC64;

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
```

```
alignment_block:
US-09-388-090-3 x YYXA_BACSU
Align seg 1/1 to: YYXA_BACSU from: 1 to: 400

718 GCCATCAATCCGGCAATTCGGCGGCCG 744
|||||
237 AlaileAsnProGlyAsnSerGlyGly 245

seq_name: SwissProt_39:YH01_HAEIN

seq_documentation_block:
ID YH01_HAEIN STANDARD; PRT; 247 AA.
AC P44292;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE HYPOTHETICAL PROTEIN H11701.
GN H11701.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN:RD / KW20;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Keiley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae RD.";
RL Science 269:496-512(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; U32843; AAC23347.1; -
CC DR TIGR; H11701; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 11 31 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
SQ SEQUENCE 247 AA; 27612 MW; 8242407A56F8D739 CRC64;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3/rev x YH01_HAEIN
Align seg 1/1 to: YH01_HAEIN from: 1 to: 247

575 GCATTGCCGATTTTGCACGACGGGT 552
|||||
118 GlyLeuProIleLeuThrGly 125

seq_name: SwissProt_39:TF2B_PYRWO

seq_documentation_block:
ID TF2B_PYRWO STANDARD; PRT; 261 AA.
```

AC P29095;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRANSCRIPTION INITIATION FACTOR IIB HOMOLOG (TFIIB).  
 GN TFB.  
 OS Pyrococcus woesei.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93324384; PubMed=8332505;  
 RA Creli R., Londei P., Cammarano P.;  
 RT "Complete nucleotide sequence of an archaeal (Pyrococcus woesei) gene  
 encoding a homolog of eukaryotic transcription factor IIB (TFIIB).";  
 RL Nucleic Acids Res. 21:2942-2942(1993).  
 RN [2]  
 RP SEQUENCE OF 110-261 FROM N.A.  
 RX MEDLINE=92130258; PubMed=1723106;  
 RA Creli R., Citarella F., Tiboni O., Sanangelantoni A.M., Palm P.,  
 RA Cammarano P.;  
 RT "Nucleotide sequence of a DNA region comprising the gene for  
 elongation factor I alpha (EF-1 alpha) from the ultrathermophilic  
 archaeote Pyrococcus woesei: phylogenetic implications.";  
 RL J. Mol. Evol. 33:332-342(1991).  
 RN [3]  
 RP SIMILARITY TO EUKARYOTIC TFIIB  
 RX MEDLINE=93046619; PubMed=1423586;  
 RA Ouzounis C., Sander C.;  
 RT "TFIIB, an evolutionary link between the transcription machineries of  
 archaeobacteria and eukaryotes.";  
 RL Cell 71:189-190(1992).  
 RN [4]  
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 62-261.  
 RX MEDLINE=97322320; PubMed=9177165;  
 RA Kosa P.F., Ghosh G., Dedecker B.S., Sigler P.B.;  
 RT "The 2.1-A crystal structure of an archaeal preinitiation complex:  
 TATA-box-binding protein/transcription factor (TII)B core/TATA-box.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:6042-6047(1997)  
 CC -1- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEL BOX-A PROMOTER.  
 CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-  
 CC INITIATION COMPLEX (DNA-TBP-TFIIB) (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE TFIIB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X70668; CAA50006.1; -  
 DR EMBL; X59857; -; NOT\_ANNOTATED\_CDS.  
 DR PIR; S31706; S31706.  
 DR PDB; 1A1S; 07-JUL-97.  
 DR INTERPRO; IPR000812; -  
 DR PFAM; PF00382; transcript\_fac2; 2.  
 DR PROSITE; PS00782; TFIIB; 2.  
 KW Transcription regulation; Repeat; 3D-structure.  
 FT REPEAT 80 156  
 FT REPEAT 176 252  
 SQ SEQUENCE 261 AA; 29748 MW; 21F1493CDEFADB64 CRC64;

alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x TF2B\_PYRWO ..

Align seg 1/1 to: TF2B\_PYRWO from: 1 to: 261

1318 CGGGTTTCGACGCGCAGACGC 1341  
 61 ArgValSerAspAlaAlaGluArg 68

seq\_name: SwissProt\_39:TF2B\_PYRFU

seq\_documentation\_block:

ID TF2B\_PYRFU STANDARD; PRT; 300 AA.  
 AC Q51731;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE TRANSCRIPTION INITIATION FACTOR IIB HOMOLOG (TFIIB).  
 GN TFB.  
 OS Pyrococcus furiosus.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zeng Q., Lewis L.M., Colangelo C.M., Dong J., Scott R.A.;  
 RT "A transcription factor (TFIIB) homolog from the hyperthermophilic  
 archaeon Pyrococcus furiosus binds Zn or Fe in an N-terminal Cys4  
 motif.";  
 RL J. Biol. Inorg. Chem. 1:162-168(1996).  
 RN [2]  
 RP STRUCTURE BY NMR OF 1-50.  
 RX MEDLINE=96163429; PubMed=8564536;  
 RA Zhu W., Zeng Q., Colangelo C.M., Lewis L.M., Summers M.F.,  
 RA Scott R.A.;  
 RT "The N-terminal domain of TFIIB from Pyrococcus furiosus forms a zinc  
 ribbon.";  
 RL Nat. Struct. Biol. 3:122-124(1996).  
 CC -1- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEL BOX-A PROMOTER.  
 CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-  
 CC INITIATION COMPLEX (DNA-TBP-TFIIB).  
 CC -1- SIMILARITY: BELONGS TO THE TFIIB FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U48391; AAC3724.1; -  
 DR PDB; 1PFT; 17-AUG-96.  
 DR INTERPRO; IPR000812; -  
 DR PFAM; PF00382; transcript\_fac2; 2.  
 DR PRINTS; PR00685; TIFACTORIIB.  
 DR PROSITE; PS00782; TFIIB; 2.  
 KW Transcription regulation; Repeat; Zinc-finger; 3D-structure.  
 FT ZN\_FING 7 29  
 SQ SEQUENCE 300 AA; 34105 MW; 476D7CA32B2ED4C1 CRC64;

alignment\_scores:

Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x TF2B\_PYRFU ..

Align seg 1/1 to: TF2B\_PYRFU from: 1 to: 300

1318 CGGGTTTCGACGCGCAGACGC 1341  
 100 ArgValSerAspAlaAlaGluArg 107

seq\_name: SwissProt\_39:TF2B\_PYRHO

seq\_documentation\_block:

```
ID ID TF2B_PYRHO STANDARD; PRT; 300 AA.
AC O59151;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE TRANSCRIPTION INITIATION FACTOR IIB HOMOLOG (TFIIB).
GN TFB OR PHI482.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE-98344137; PubMed-9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RT DNA Res. 5:55-76(1998).
CC -!- FUNCTION: STABILIZES TBP BINDING TO AN ARCHAEAL BOX-A PROMOTER.
CC ALSO RESPONSIBLE FOR RECRUITING RNA POLYMERASE II TO THE PRE-
CC INITIATION COMPLEX (DNA-TBP-TFIIB) (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TFIIB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AP000005; BAA30589.1; -
DR INTERPRO; IPR000812; -
DR PFAM; PF00382; transcript_fac2; 2.
DR PRINTS; PR00685; TIFACTORIIB.
DR PROSITE; PS00782; TFIIB; 2.
KW Transcription regulation; Repeat; Zinc-finger.
FT ZN_FING 7
FT SEQUENCE 300 AA; 34097 MW; DE9758F398BC855F CRC64;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x TF2B_PYRHO ..
Align seg 1/1 to: TF2B_PYRHO from: 1 to: 300

1318 CGGGTTCCGACCGCGCAGAACCC 1341
|||||
100 ArgValSerAspAlaAlaGluArg 107

seq_name: SwissProt_39:CYCR_RHOGE

seq_documentation_block:
ID CYCR_RHOGE STANDARD; PRT; 366 AA.
AC P51758;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHOTOSYNTHETIC REACTION CENTER CYTOCHROME C SUBUNIT PRECURSOR.
GN PUF.
OS Rhodococcus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Rubrivivax.
RN [1]
```

```
RP SEQUENCE FROM N.A.
RX STRAIN=111144;
RA MEDLINE-94132007; PubMed-8300574;
RA Nagashima K.V.P., Matsuura K., Ohshima S., Shimada K.;
RT "Primary structure and transcription of genes encoding B870 and
RT photosynthetic reaction center apoproteins from Rubrivivax
RT gelatinosus.";
RL J. Biol. Chem. 269:2477-2484(1994).
CC -!- FUNCTION: THE REACTION CENTER OF PURPLE BACTERIA CONTAIN A
CC TIGHTLY BOUND CYTOCHROME MOLECULE WHICH RE-REDUCES THE PHOTO
CC OXIDIZED PRIMARY ELECTRON DONOR.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A LIPID
CC ANCHOR (BY SIMILARITY).
CC -!- PTM: BINDS FOUR HEME GROUPS PER MOLECULE.
CC -!- SIMILARITY: HIGH, WITH OTHER PHOTOSYNTHETIC REACTION CENTER
CC CYTOCHROME C SUBUNITS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D16822; BAA04102.1; -
DR HSSP; P07173; 4PRC.
DR INTERPRO; IPR000345; -
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
DR PROSITE; PS00190; CYTOCHROME_C; 4.
KW Electron transport; Photosynthesis; Reaction center; Heme;
KW Membrane; Lipoprotein; Duplication; Signal.
FT SIGNAL 1 22
FT CHAIN 23 366
FT -----
FT C SUBUNIT
FT PHOTOSYNTHETIC REACTION CENTER CYTOCHROME
FT LIPID 23 23
FT BINDING 107 107
FT BINDING 110 110
FT METAL 111 111
FT -----
FT BINDING 151 151
FT BINDING 154 154
FT METAL 155 155
FT -----
FT BINDING 249 249
FT BINDING 252 252
FT METAL 253 253
FT -----
FT BINDING 309 309
FT BINDING 312 312
FT METAL 313 313
FT SEQUENCE 366 AA; 39195 MW; 2480B13766316031 CRC64;

alignment_scores:
  Quality: 8.00 Length: 8
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x CYCR_RHOGE ..
Align seg 1/1 to: CYCR_RHOGE from: 1 to: 366

43 GCCGCCTTCTGCGCAGCTGCCAA 66
|||||
17 AlaAlaLeuAlaAlaGlyCysGlu 24

seq_name: SwissProt_39:G154_MOUSE
seq_documentation_block:
ID G154_MOUSE STANDARD; PRT; 429 AA.
```

AC P50636;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE GAMETOGENESIS EXPRESSED PROTEIN GEG-154.  
 GN GEG-154.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SWISS; TISSUE=OVARY;  
 RX MEDLINE=95341612; PubMed=7616504;  
 RA Lopez-Alanor D.M., del Mazo J.;  
 RT "Cloning and characterization of genes expressed during gametogenesis  
 of female and male mice."  
 RL J. Reprod. Fert. 103:323-329(1995).  
 CC -1- DEVELOPMENTAL STAGE: PREFERENTIALLY EXPRESSED IN BOTH SEXES DURING  
 CC GAMETOGENESIS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X71642; CAA50643.1; -  
 KW Transmembrane.  
 FT TRANSMEM 39 59 POTENTIAL.  
 FT TRANSMEM 154 174 POTENTIAL.  
 SQ SEQUENCE 429 AA; 45213 MW; 7A3132693379C60F CRC64;  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-388-090-3/rev x G154\_MOUSE ..  
 Align seg 1/1 to: G154\_MOUSE from: 1 to: 429  
 293 GGGTCGCTGTCGGCAAGCGGTGC 270  
 ||||||||||||||||||||  
 128 GlySerLeuSerAlaSerGlySer 135  
 seq\_name: SwissProt\_39:YIEO\_HAEIN  
 seq\_documentation\_block:  
 ID YIEO\_HAEIN STANDARD; PRT; 463 AA.  
 AC P44903;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DE 01-NOV-1995 (Rel. 32, Last annotation update)  
 DE HYPOTHETICAL PROTEIN HI0852.  
 GN HI0852.  
 OS Haemophilus influenzae.  
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
 OC Haemophilus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RD / KW20;  
 RX MEDLINE=95350630; PubMed=7542800;  
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
 RA McInney K., Sutton G., Smith R., Fields C.A., Gocayne J.D.,  
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,

RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
 RA Venter J.C.;  
 RT "Whole-genome random sequencing and assembly of Haemophilus  
 RT influenzae Rd."  
 RL Science 269:496-512(1995).  
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE  
 CC (PROBALE).  
 CC -1- SIMILARITY: STRONG, TO E.COLI YIEO.  
 CC -1- SIMILARITY: BELONGS TO THE MAJOR FACILITATOR FAMILY (ALSO KNOWN  
 CC AS THE DRUG RESISTANCE TRANSLOCASE FAMILY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; U32766; AAC22509.1; -  
 DR TIGR; HI0852; -  
 KW Hypothetical protein; Transport; Transmembrane; Inner membrane.  
 FT TRANSMEM 10 30 POTENTIAL.  
 FT TRANSMEM 49 69 POTENTIAL.  
 FT TRANSMEM 82 102 POTENTIAL.  
 FT TRANSMEM 107 127 POTENTIAL.  
 FT TRANSMEM 139 159 POTENTIAL.  
 FT TRANSMEM 165 185 POTENTIAL.  
 FT TRANSMEM 197 217 POTENTIAL.  
 FT TRANSMEM 225 245 POTENTIAL.  
 FT TRANSMEM 267 287 POTENTIAL.  
 FT TRANSMEM 298 318 POTENTIAL.  
 FT TRANSMEM 328 348 POTENTIAL.  
 FT TRANSMEM 354 374 POTENTIAL.  
 FT TRANSMEM 393 413 POTENTIAL.  
 FT TRANSMEM 429 449 POTENTIAL.  
 SQ SEQUENCE 463 AA; 50002 MW; D725EFFD50F9F23F CRC64;  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 alignment\_block:  
 US-09-388-090-3/rev x YIEO\_HAEIN ..  
 Align seg 1/1 to: YIEO\_HAEIN from: 1 to: 463  
 273 GTCGAATCGGTTTCGGCATGTC 250  
 ||||||||||||||||||||  
 400 ValGlyIleGlyPheGlyIleAla 407  
 seq\_name: SwissProt\_39:SYE\_BACST  
 seq\_documentation\_block:  
 ID SYE\_BACST STANDARD; PRT; 489 AA.  
 AC P22249;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 01-JUN-1994 (Rel. 29, Last annotation update)  
 DE GLUTAMYL--TRNA SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE)  
 DE (GLURS).  
 GN GLURS.  
 OS Bacillus stearothermophilus.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;  
 OC Bacillus/Staphylococcus group; Bacillus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91009314; PubMed=2120226;  
 RA Berton R., Watson D., Yaguchi M., Lapointe J.;  
 RA "GlutamyI-trna synthetases of Bacillus subtilis 168T and of Bacillus  
 RT stearothermophilus. Cloning and sequencing of the glx genes and

RT comparison with other aminoacyl-tRNA synthetases. ;  
 RL J. Biol. Chem. 265:18248-18255(1990).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) ~ AMP +  
 CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).  
 CC -!- SUBUNIT: MONOMER.  
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.  
 CC -!- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; M5072; AAA22494.1; -;  
 DR PIR; B36090; SYBSES.  
 DR HSP; P27000; IGLN.  
 DR INTERPRO; IPR000924; -;  
 DR INTERPRO; IPR001412; -;  
 DR PFAM; PF00749; trna-synt\_1c; 1.  
 DR PRINTS; PS00987; TRNASYNTHGLU.  
 DR PROSITE; PS00178; AA-TRNA-LIGASE\_I; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.  
 FT SIMILAR 11 21 "HIGH" REGION.  
 FT SIMILAR 253 257 "KMSKS" REGION.  
 FT BINDING 256 256 ATP (BY SIMILARITY).  
 SQ SEQUENCE 489 AA; 56183 MW; AB71266CB80A2E15 CRC64;  
 -----  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 -----  
 alignment\_block:  
 US-09-388-090-3/rev x SYE\_BACST ..  
 Align seg 1/1 to: SYE\_BACST from: 1 to: 489  
 753 GAACAGCGCGCGCGGATTCGCC 730  
 |||||||||||||||||||||  
 121 GUGlnArgAlaGlyIleAla 128  
 -----  
 seq\_name: SwissProt\_39:DEGP\_BARHE  
 seq\_documentation\_block:  
 ID DEGP\_BARHE STANDARD; PRT; 503 AA.  
 AC P54925;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE PROBABLE PERIPLASMIC SERINE PROTEASE DO-LIKE PRECURSOR (EC 3.4.21.-)  
 DE (HTRA ANTIGEN).  
 GN DEGP OR HTRA.  
 OS Bartonella henselae (Rochalimaea henselae).  
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;  
 OC Bartonellaceae; Bartonella.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-HOUSTON-1.  
 RX MEDLINE-94299828; PubMed-8027347;  
 RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,  
 RA Goral S., Hager C., Edwards K.;  
 RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch  
 RT disease patients by PCR.";  
 RL J. Clin. Microbiol. 32:942-948(1994).  
 CC -!- SUBCELLULAR LOCATION: PERIPLASMIC (POTENTIAL).  
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE  
 CC DEGP/DEGO/DEGS FAMILY.  
 CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.  
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL; L20127; AAA97430.1; -;  
 DR MEROPS; S01\_273; -;  
 DR INTERPRO; IPR001254; -;  
 DR INTERPRO; IPR001478; -;  
 DR INTERPRO; IPR001940; -;  
 DR PFAM; PF00595; PDZ; 2.  
 DR PFAM; PF00089; trypsin; 1.  
 DR PRINTS; PR00834; PROTEASES2C.  
 KW Hydrolase; Serine protease; Periplasmic; Signal.  
 FT SIGNAL 1 18  
 FT CHAIN 19 503  
 FT LIKE.  
 FT ACT\_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).  
 FT ACT\_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).  
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;  
 -----  
 alignment\_scores:  
 Quality: 8.00 Length: 8  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000  
 -----  
 alignment\_block:  
 US-09-388-090-3 x DEGP\_BARHE ..  
 Align seg 1/1 to: DEGP\_BARHE from: 1 to: 503  
 640 GTGACCGCGCGCATCGTGTCGCC 663  
 |||||||||||||||||||||  
 215 ValThrAlaGlyIleValSerAla 222  
 -----  
 seq\_name: SwissProt\_39:FXC1\_HUMAN  
 seq\_documentation\_block:  
 ID FXC1\_HUMAN STANDARD; PRT; 553 AA.  
 AC Q12948; Q9UP06;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-  
 DE RELATED TRANSCRIPTION FACTOR 3) (FREAC-3).  
 GN FOXC1 OR FKHL7 OR FREAC3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND VARIANTS S-112; M-126 AND L-131.  
 RX MEDLINE-98282091; PubMed-9620769;  
 RA Nishimura D.Y., Swiderski R.E., Alward W.L.M., Searby C.C.,  
 RA Patil S.R., Bennet S.R., Kanis A.B., Gastier J.M., Stone E.M.,  
 RA Sheffield V.C.;  
 RA "The forkhead transcription factor gene FKHL7 is responsible for  
 RT glaucoma phenotypes which map to 6p25.";  
 RL Nat. Genet. 19:140-147(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND VARIANTS ARA THR-82 AND MET-87.  
 RX MEDLINE-99011252; PubMed-9792859;  
 RA Mears A.J., Jordan T., Mirzayans F., Dubois S., Kume T., Parlee M.,  
 RA Ritch R., Koop B., Kuo W.-L., Collins C., Marshall J., Gould D.B.,  
 RA Pearce W., Carlsson P., Enerbaeck S., Morissette J., Bhattacharya S.,  
 RA Hogan B., Raymond V., Walter M.A.;  
 RT "Mutations of the forkhead/winged-helix gene, FKHL7, in patients with  
 RT Axenfeld-Rieger anomaly.";  
 RL Am. J. Hum. Genet. 63:1316-1328(1998).

```
[3]
RN SEQUENCE OF 73-178 FROM N.A.
RX MEDLINE=95045392; PubMed=7957066;
RA Pierrou S., Helliqvist M., Samuelsson L., Enerbaeck S., Carlsson P.;
RT "Cloning and characterization of seven human forkhead proteins:
RL binding site specificity and DNA bending.";
RL EMBO J. 13:5002-5012(1994).
CC -!- FUNCTION: BINDING OF FREAC-3 AND FREAC-4 TO THEIR COGNATE SITES
CC RESULTS IN BENDING OF THE DNA AT AN ANGLE OF 80-90 DEGREES.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- DISEASE: DEFECTS IN FKHL7 ARE THE CAUSE A SPECTRUM OF GLAUCOMA
CC PHENOTYPES SUCH AS AXENFELD-RIEGER ANOMALY (ARA), AXENFELD-RIEGER
CC SYNDROME (ARS) AND IRIDOGONIOSGENESIS ANOMALY (IGDA). ARS IS AN
CC AUTOSOMAL DOMINANT DISORDER PRESENTING WITH ARA-LIKE OCULAR
CC FINDINGS IN ADDITION TO ABNORMALITIES OF THE TEETH, JAW AND
CC UMBILICUS.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF048693; AAC18081.1; -.
DR EMBL; U13221; AAA92038.1; -.
DR EMBL; AF078096; AAC72915.1; -.
DR MIM; 601090; -.
DR MIM; 601631; -.
DR INTERPRO; IPR001766; -.
DR INTERPRO; IPR002952; -.
DR INTERPRO; IPR002965; -.
DR PFAM; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR PRINTS; PR01217; PRICHEXTENS.
DR PRINTS; PR01228; EGGSHELL.
DR PROSITE; PS00657; FORK_HEAD_1; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
KW DNA-binding; Nuclear protein; Transcription regulation;
Disease mutation.
FT DNA_BIND 77 168
FT DOMAIN 28 33
FT DOMAIN 169 173
FT DOMAIN 194 197
FT DOMAIN 262 272
FT DOMAIN 292 297
FT DOMAIN 375 382
FT DOMAIN 438 445
FT DOMAIN 447 456
FT DOMAIN 486 495
FT VARIANT 82 82
FT VARIANT 87 87
FT VARIANT 112 112
FT VARIANT 126 126
FT VARIANT 131 131
FT CONFLICT 180 180
FT CONFLICT 199 202
FT CONFLICT 426 426
SQ SEQUENCE 553 AA; 56787 MW; DBC81B943030359E CRC64;
```

alignment\_scores:

Quality: 8.00  
Ratio: 1.000Length: 8  
Gaps: 0

```
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-388-090-3 x FXCL_HUMAN ..
Align seq 1/1 to: FXCL_HUMAN from: 1 to: 553
599 GGCTCGCTGCATCGCGCGCCCT 622
|||||
275 GlySerLeuProSerAlaArgPro 282
seq_name: SwissProt_39:FXCL_MOUSE
seq_documentation_block:
ID FXCL_MOUSE STANDARD; PRT; 553 AA.
AC Q61572; Q61582; O88409;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FORKHEAD BOX PROTEIN C1 (FORKHEAD-RELATED PROTEIN FKHL7) (FORKHEAD-
DE RELATED TRANSCRIPTION FACTOR 3) (FREAC-3) (TRANSCRIPTION FACTOR FKHL-1)
DE (MESODERM/MESENCHYME FORKHEAD 1) (MF-1).
GN FOXC1 OR FKHL7 OR FREAC3 OR FKHL1 OR MFL.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98297351; PubMed=9635428;
RA Kume T., Deng K.Y., Winfrey V., Gould D.B., Walter M.A., Hogan B.L.M.;
RT "The forkhead/winged helix gene Mfl is disrupted in the pleiotropic
RL Cell 93:985-996(1998).
RN [2]
RP SEQUENCE OF 69-179 FROM N.A.
RC STRAIN=129;
RX MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
RT regulated.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
RN [3]
RP SEQUENCE OF 71-187 FROM N.A.
RX MEDLINE=93387221; PubMed=8375339;
RA Sasaki H., Hogan B.L.;
RT "Differential expression of multiple fork head related genes during
RT gastrulation and axial pattern formation in the mouse embryo.";
RL Development 118:47-59(1993).
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN MANY EMBRYONIC TISSUES, INCLUDING
CC PRECHONDROGENIC MESENCHYME, PERIOULAR MESENCHYME, MENINGES,
CC ENDOTHELIAL CELLS AND KIDNEY.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -!- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF045017; AAC24209.1; -.
DR EMBL; L10406; AAA03159.1; -.
DR EMBL; X71939; CAA50741.1; -.
DR HSSP; Q63245; 2HFH.
DR TRANSFAC; T02426; -.
DR MGD; MGI:1347466; FOXC1.
DR INTERPRO; IPR001766; -.
DR PFAM; PF00250; Fork_head; 1.
DR PRINTS; PR00353; FORKHEAD.
```

DR PROSITE; PS00657; FORK\_HEAD\_1; 1.  
DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
DR PROSITE; PS50039; FORK\_HEAD\_3; 1.  
KW DNA-binding; Nuclear protein; Transcription regulation.  
FT DOMAIN 28 33 POLY-ALA.  
FT DNA\_BIND 77 168 FORK-HEAD.  
FT DOMAIN 169 173 POLY-ARG.  
FT DOMAIN 194 197 POLY-PRO.  
FT DOMAIN 264 274 POLY-SER.  
FT DOMAIN 375 386 POLY-GLY.  
FT DOMAIN 444 451 POLY-SER.  
FT DOMAIN 453 456 POLY-GLY.  
FT DOMAIN 486 496 POLY-ALA.  
FT CONFLICT 180 187 VKDKEEKG -> KKEITFIG (IN REF. 3).  
SQ SEQUENCE 553 AA; 56953 MW; 3CDD12F69CA4F217 CRC64;

alignment\_scores:  
Quality: 8.00 Length: 8  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x FXCL\_MOUSE ..  
Align seg 1/1 to: FXCL\_MOUSE from: 1 to: 553  
599 GGGTCGCTGCCATCGGCGGCCCT 622  
|||||  
277 GlySerLeuProSerAlaArgPro 284



OM of: US-09-388-090-3 to: SPTREMBL15:\* out\_format : pfs

Date: Mar 22, 2001 12:08 PM

About: Results were produced by the GenCore software, version 4.5,  
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09388090/runat\_20032001\_092951\_25846/app\_query.fasta\_1.1474  
-DB=SPTREMBL15 -OPMT=fastan -SUFFIX=oligo7.rspt -GAPOP=4.500  
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -GAPOP=60.000 -LOOPEXT=0.000  
-GAPOP=6.000 -GAPEXT=7.000 -GAPOP=60.000 -GAPEXT=60.000  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo  
-TRANS=human4.cdi -LST=1000 -DOCALLIGN=200 -THR\_SCORE=quality  
-THR\_MIN=7 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09388090\_@CGN1\_1\_123 -NCPU=6 -ICPU=3  
-LONGLOG -NO\_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-388-090-3

Query length: 1395

Database: SPTREMBL15:\*

Database sequences: 374700

Database length: 117207915

Search time (sec): 62.980000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000

WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

score\_list:

Sequence	Strd	Orig	ZScore	E5Score	Len	Documentation
sp_bacteria:Q9JVT1	+	127.00	2346.04	3.4e-123	499	Q9JVT1 neisseria meningitidis
sp_bacteria:Q44476	+	19.00	324.28	1.5e-10	473	Q44476 azotobacter vinelandii
sp_bacteria:Q57155	+	19.00	324.27	1.5e-10	474	Q57155 pseudomonas aeruginosa
sp_bacteria:Q35021	+	13.00	211.59	0.0003	500	Q35021 bacillus subtilis
sp_bacteria:Q72780	+	11.00	179.25	0.0371	224	Q72780 myxococcus xanthus
sp_bacteria:Q05670	+	11.00	175.66	0.0335	394	Q05670 bacillus subtilis
sp_bacteria:Q83752	+	11.00	175.61	0.0334	397	Q83752 streptococcus pneumoniae
sp_plant:Q04674	+	11.00	175.59	0.0334	398	Q04674 haematomia pallidum
sp_bacteria:Q73940	+	11.00	175.31	0.0331	416	Q73940 synecocystis sp. (strai)
sp_plant:Q9LK85	+	11.00	174.97	0.0328	439	Q9LK85 arabidopsis thaliana
sp_plant:Q9JTK4	+	11.00	174.94	0.0328	441	Q9JTK4 gainococcus radiodurans
sp_plant:Q9JUI0	+	11.00	174.84	0.0327	448	Q9JUI0 arabidopsis thaliana
sp_bacteria:Q34358	+	11.00	174.83	0.0327	449	Q34358 bacillus subtilis
sp_bacteria:Q73354	+	11.00	174.78	0.0326	452	Q73354 synecocystis sp. (strai)
sp_bacteria:Q67436	+	11.00	174.77	0.0326	453	Q67436 aquifex aeolicus
sp_bacteria:Q9R911	+	11.00	174.70	0.0326	458	Q9R911 bacillus subtilis
sp_bacteria:Q9K1W4	+	11.00	174.69	0.0325	459	Q9K1W4 thermotoga maritima
sp_bacteria:Q9JTD7	+	11.00	174.30	0.0322	488	Q9JTD7 chlamydia pneumoniae
sp_bacteria:Q9PL97	+	11.00	174.30	0.0322	488	Q9PL97 chlamydia pneumoniae
sp_plant:Q04481	+	10.00	174.18	0.0321	497	Q04481 chlamydia muridarum
sp_plant:Q9SH20	+	10.00	156.09	0.3605	450	Q9SH20 arabidopsis thaliana
sp_plant:Q82261	+	10.00	155.19	0.3512	518	Q82261 arabidopsis thaliana
sp_bacteria:Q86761	+	9.00	140.05	0.3386	632	Q86761 streptomyces coelicolor
sp_bacteria:Q9S285	+	9.00	138.75	4.14	362	Q9S285 streptomyces coelicolor
sp_bacteria:Q51374	+	9.00	138.29	4.09	389	Q51374 pseudomonas aeruginosa
sp_bacteria:Q9K5R6	+	9.00	138.02	4.05	406	Q9K5R6 bacillus halodurans
sp_bacteria:Q9Z4H7	+	9.00	137.99	4.05	408	Q9Z4H7 lactococcus lactis
sp_invertebrate:Q9VJF3	+	9.00	137.77	4.03	422	Q9VJF3 drosophila melanogaster
sp_bacteria:Q9Z566	+	9.00	137.34	3.98	452	Q9Z566 mycobacterium leprae
sp_bacteria:Q53896	+	9.00	137.17	3.96	464	Q53896 mycobacterium tuberculosis
sp_bacteria:Q44596	+	9.00	137.03	3.94	474	Q44596 bruceella abortus
sp_bacteria:Q9PBA3	+	9.00	136.52	3.88	514	Q9PBA3 xylella fastidiosa
sp_plant:Q80829	+	9.00	134.37	3.65	720	Q80829 arabidopsis thaliana
sp_invertebrate:Q96641	+	9.00	134.23	3.63	737	Q96641 drosophila virilis
sp_plant:Q80394	+	9.00	133.78	3.59	790	Q80394 mesembryanthemum crystal
sp_invertebrate:Q00901	-	9.00	133.53	3.56	822	Q00901 leishmania major

sp_plant:Q9ZPU5	-	9.00	130.45	3.26	1335	Q9ZPU5 arabidopsis thaliana
sp_plant:Q9SH77	-	9.00	130.35	3.25	1356	Q9SH77 arabidopsis thaliana
sp_plant:Q9M819	+	9.00	128.86	3.11	1713	Q9M819 arabidopsis thaliana
sp_invertebrate:Q9NDY7	+	9.00	128.20	3.05	1899	Q9NDY7 leishmania major
sp_bacteria:Q9I946	+	8.00	124.23	51.62	187	Q9I946 pseudomonas putida
sp_invertebrate:Q9VFS8	+	8.00	123.20	50.10	220	Q9VFS8 drosophila melanogaster
sp_bacteria:Q52255	+	8.00	123.11	49.97	223	Q52255 pseudomonas coelicolor
sp_bacteria:Q9S2P8	+	8.00	121.62	47.86	282	Q9S2P8 streptomyces coelicolor
sp_human:Q9S517	+	8.00	121.44	47.61	290	Q9S517 homo sapiens (human)
sp_invertebrate:Q93727	-	8.00	121.26	47.38	298	Q93727 caenorhabditis elegans
sp_archaea:Q9V0V5	+	8.00	121.22	47.32	300	Q9V0V5 pyrococcus abyssi
sp_bacteria:Q9S0K5	+	8.00	121.05	47.09	308	Q9S0K5 shewanella violacea
sp_archaea:Q9YDE9	+	8.00	120.99	47.00	311	Q9YDE9 aeropyrum pernix
sp_bacteria:Q9ZEO2	+	8.00	120.08	45.78	339	Q9ZEO2 rickettsia prowazekii
sp_bacteria:Q9S617	+	8.00	119.96	45.62	366	Q9S617 rhodocyclus gelatinosus
sp_bacteria:Q9JPB7	+	8.00	119.96	45.62	366	Q9JPB7 rhodocyclus gelatinosus
sp_bacteria:Q46094	+	8.00	119.92	45.57	368	Q46094 campylobacter jejuni
sp_bacteria:Q31388	+	8.00	119.87	45.50	371	Q31388 bradyrhizobium japonicum
sp_plant:Q9ZPG7	+	8.00	119.52	45.04	392	Q9ZPG7 arabidopsis thaliana
sp_bacteria:Q9K6X6	+	8.00	119.42	44.92	398	Q9K6X6 bacillus halodurans
sp_human:Q9Y4Y1	+	8.00	119.34	44.82	403	Q9Y4Y1 homo sapiens (human)
sp_bacteria:Q25663	+	8.00	118.74	44.04	443	Q25663 helicobacter pylori
sp_invertebrate:Q44021	-	8.00	118.60	43.86	453	Q44021 plasmodium falciparum
sp_human:Q9NV58	+	8.00	118.54	43.79	457	Q9NV58 homo sapiens (human)
sp_bacteria:Q06439	+	8.00	118.50	43.74	460	Q06439 rhodobacter capsulatus
sp_bacteria:Q46120	+	8.00	118.34	43.53	472	Q46120 campylobacter jejuni
sp_bacteria:Q9PN69	+	8.00	118.34	43.53	472	Q9PN69 campylobacter jejuni
sp_bacteria:Q9ZM18	+	8.00	118.28	43.46	476	Q9ZM18 helicobacter pylori
sp_bacteria:Q9PGL3	+	8.00	118.22	43.38	481	Q9PGL3 xylella fastidiosa
sp_invertebrate:Q9VFM0	-	8.00	117.92	43.01	504	Q9VFM0 drosophila melanogaster
sp_plant:Q9SRL6	+	8.00	117.72	42.76	520	Q9SRL6 arabidopsis thaliana
sp_human:Q9NUE5	+	8.00	117.33	42.28	553	Q9NUE5 homo sapiens (human)
sp_rhodent:Q9OWR9	+	8.00	117.33	42.28	553	Q9OWR9 mus musculus (mouse)
sp_fungi:Q9S954	+	8.00	117.25	42.18	560	Q9S954 emericella nidulans
sp_invertebrate:Q93399	+	8.00	117.24	42.17	561	Q93399 xenopus laevis
sp_bacteria:Q9X870	+	8.00	117.05	41.94	578	Q9X870 streptomyces coelicolor
sp_invertebrate:Q9W324	-	8.00	117.05	41.94	578	Q9W324 drosophila melanogaster
sp_rhodent:Q9SAA2	+	8.00	116.71	41.53	610	Q9SAA2 mus musculus (mouse)
sp_plant:Q9SA69	+	8.00	116.46	41.23	634	Q9SA69 arabidopsis thaliana
sp_vertebrate:Q42182	+	8.00	116.01	40.69	681	Q42182 brachydanio rerio
sp_fungi:Q906825	+	8.00	115.51	40.12	736	Q906825 saccharomyces cerevisiae
sp_invertebrate:Q9V515	-	8.00	115.43	40.02	746	Q9V515 drosophila melanogaster
sp_plant:Q9X9E7	+	8.00	115.16	39.71	778	Q9X9E7 arabidopsis thaliana
sp_bacteria:Q9P6323	+	8.00	114.93	39.44	807	Q9P6323 escherichia coli
sp_rhodent:Q9OUJ5	+	8.00	114.67	39.15	840	Q9OUJ5 mus musculus (mouse)
sp_bacteria:Q87943	+	8.00	114.51	38.97	861	Q87943 thauera aromatica
sp_plant:Q9LYT4	+	8.00	112.89	37.18	1112	Q9LYT4 arabidopsis thaliana
sp_virus:Q93016	+	8.00	112.63	36.91	1157	Q93016 sarcophthine herpes
sp_plant:Q9MLR5	+	8.00	112.01	36.25	1276	Q9MLR5 saccharomyces cerevisiae
sp_bacteria:Q9MIR5	+	8.00	111.59	35.82	1363	Q9MIR5 arabidopsis thaliana
sp_bacteria:Q9S2Y9	+	8.00	110.92	35.13	1514	Q9S2Y9 streptomyces coelicolor
sp_invertebrate:Q17163	+	8.00	109.81	34.02	1802	Q17163 brugia malayi
sp_fungi:Q9Y9A2	+	8.00	107.47	31.79	2607	Q9Y9A2 gibberella fujikuroi
sp_rhodent:Q9WTS1	+	7.00	117.59	807.97	28	Q9WTS1 rattus norvegicus
sp_rhodent:Q9JLR6	+	7.00	117.36	802.77	29	Q9JLR6 rattus norvegicus
sp_virus:Q90088	+	7.00	116.17	775.47	35	Q90088 human papillomavirus
sp_virus:Q9YML4	+	7.00	116.17	775.47	35	Q9YML4 human immunodeficiency
sp_bacteria:Q9K2L1	+	7.00	115.16	753.22	41	Q9K2L1 vibrio cholerae
sp_virus:Q84308	+	7.00	114.71	743.50	44	Q84308 human papillomavirus
sp_bacteria:Q9RR49	+	7.00	113.41	716.01	54	Q9RR49 azotobacter chroococcum
sp_bacteria:Q9KMH3	+	7.00	112.85	704.44	59	Q9KMH3 vibrio cholerae
sp_bacteria:Q9KME8	+	7.00	112.85	704.44	59	Q9KME8 vibrio cholerae
sp_bacteria:Q9KMC7	+	7.00	112.85	704.44	59	Q9KMC7 vibrio cholerae
sp_bacteria:Q9KMF3	+	7.00	112.74	702.26	60	Q9KMF3 vibrio cholerae
sp_bacteria:Q9JWM2	+	7.00	111.67	680.85	71	Q9JWM2 neisseria meningitidis
sp_plant:Q9LFR23	+	7.00	111.67	680.85	71	Q9LFR23 arabidopsis thaliana
sp_phage:Q9XJ55	+	7.00	111.40	675.68	74	Q9XJ55 bacteriophage pm2
sp_bacteria:Q9JXC0	+	7.00	110.91	666.06	80	Q9JXC0 neisseria meningitidis
sp_virus:Q72946	+	7.00	110.60	660.11	84	Q72946 hepatitis c virus
sp_bacteria:Q9RUX4	+	7.00	110.37	655.86	87	Q9RUX4 deinococcus radiodurans
sp_bacteria:Q9Z384	+	7.00	110.30	654.48	88	Q9Z384 actinobacillus actinom
sp_phage:Q38613	+	7.00	110.30	654.48	88	Q38613 bacteriophage tuc2009
sp_bacteria:Q54156	+	7.00	109.95	647.86	93	Q54156 shigella flexneri

sp_virus:O31294	+	7.00	109.88	646.59	94	! 091294 human immunodeficiency vi
sp_virus:O90727	+	7.00	109.68	642.86	97	! 090727 human papillomavirus type
sp_virus:Q1G75	+	7.00	109.68	642.86	97	! 090727 porcine adenovirus 3. 97
sp_vertibrate:Q9PSD5	+	7.00	109.43	638.10	101	! 09psd5 gallus gallus (chicken).
sp_bacteria:O51757	+	7.00	109.24	634.67	104	! 051757 pseudomonas fluorescens.
sp_bacteria:O26041	+	7.00	109.00	630.28	108	! 026041 helicobacter pylori (can
sp_bacteria:P94670	+	7.00	109.00	630.28	108	! 094670 fremyella diplosiphon (o
sp_invertebrate:Q9VAA6	+	7.00	108.82	627.11	111	! 09vaa6 drosophila melanogaste
sp_virus:Q9QRC2	+	7.00	108.77	626.08	112	! 09qrc2 human immunodeficiency v
sp_bacteria:O69507	+	7.00	108.54	622.05	116	! 069507 mycobacterium leprae. hy
sp_bacteria:Q9KTG0	+	7.00	108.49	621.07	117	! 09ktg0 vibrio cholerae. hypothe
sp_bacteria:Q9Y9V3	+	7.00	108.38	619.13	119	! 09y9v3 aeropyrum pernix. hypoth
sp_bacteria:Q9X8E5	+	7.00	108.22	616.30	122	! 09x8e5 streptomyces coelicolor.
sp_archaea:Q9YAB3	+	7.00	108.12	614.46	124	! 09yab3 aeropyrum pernix. hypoth
sp_virus:Q9503	+	7.00	107.97	611.77	127	! 09503 garlic virus b. nucleic
sp_archaea:Q9Y912	+	7.00	107.77	608.29	131	! 09y912 aeropyrum pernix. hypoth
sp_plant:Q9SUX3	+	7.00	107.67	606.59	133	! 09sux3 arabidopsis thaliana (mc
sp_bacteria:O31429	+	7.00	107.49	603.30	137	! 031429 bacillus subtilis. ybdo
sp_archaea:Q9YE28	+	7.00	107.44	602.49	138	! 09ye28 aeropyrum pernix. hypoth
sp_bacteria:P73805	+	7.00	107.35	600.90	140	! 073805 synechocystis sp. (strai
sp_virus:Q9J7G8	+	7.00	107.35	600.90	140	! 09j7g8 coxsackievirus b4. polyt
sp_bacteria:O66357	+	7.00	107.26	599.33	142	! 066357 unidentified nitrogen-fi
sp_mammal:O28568	+	7.00	107.12	597.03	145	! 028568 ovis aries (sheep). earl
sp_plant:O4166	+	7.00	107.12	597.03	145	! 04166 physcomitrella patens (m
sp_virus:O73443	+	7.00	107.12	597.03	145	! 073443 human papillomavirus. la
sp_bacteria:Q43956	+	7.00	106.95	594.05	149	! 043956 azotobacter chroococcum
sp_bacteria:Q9X549	+	7.00	106.95	594.05	149	! 09x549 corynebacterium glutamic
sp_bacteria:Q9L266	+	7.00	106.78	591.16	153	! 09l266 streptomyces coelicolor
sp_invertebrate:Q17550	+	7.00	106.78	591.16	153	! 017550 caenorhabditis elegans
sp_mammal:Q9XS44	+	7.00	106.78	591.16	153	! 09xs44 tayassu tajacu (collared
sp_virus:Q81071	+	7.00	106.78	591.16	153	! 081071 human papillomavirus. m
sp_virus:Q84228	+	7.00	106.78	591.16	153	! 084228 human papillomavirus. la
sp_virus:Q81072	+	7.00	106.78	591.16	153	! 081072 human papillomavirus. m
sp_virus:O81073	+	7.00	106.78	591.16	153	! 081073 human papillomavirus. m
sp_plant:O49945	+	7.00	106.66	589.05	156	! 049945 solanum tuberosum (potat
sp_mammal:Q97659	+	7.00	106.66	589.05	156	! 097659 oryctolagus cuniculus (r
sp_invertebrate:Q9NSV2	+	7.00	106.08	579.19	171	! 09nsv2 caenorhabditis elegans
sp_invertebrate:Q9Z1Y2	+	7.00	106.00	577.95	173	! 09z1y2 cavia porcellus (guinea
sp_invertebrate:Q9N474	+	7.00	105.96	577.34	174	! 09n474 caenorhabditis elegans
sp_virus:Q9QEK8	+	7.00	105.96	577.34	174	! 09qek8 canine distemper virus
sp_vertibrate:P70069	-	7.00	105.93	576.73	175	! 070069 gallus gallus (chicken).
sp_human:Q9PLR8	+	7.00	105.89	576.12	176	! 09plr8 homo sapiens (human). ho
sp_bacteria:O84363	+	7.00	105.82	574.93	178	! 084363 chlamydia trachomatis. h
sp_human:Q9UF32	+	7.00	105.75	573.75	180	! 09uf32 homo sapiens (human). h
sp_fungi:Q9USE4	+	7.00	105.71	573.16	181	! 09use4 schizosaccharomyces pom
sp_invertebrate:Q9VSB3	-	7.00	105.71	573.16	181	! 09vsb3 drosophila melanogaste
sp_plant:Q9M054	+	7.00	105.64	572.00	183	! 09m054 arabidopsis thaliana (mc
sp_virus:Q9IGT3	+	7.00	105.61	571.43	184	! 09igt3 porcine adenovirus 3. 18
sp_bacteria:Q9K394	+	7.00	105.51	569.73	187	! 09k394 rhodobacter capsulatus
sp_bacteria:Q9KYR3	+	7.00	105.51	569.73	187	! 09kyr3 streptomyces coelicolor.
sp_invertebrate:Q9VD16	+	7.00	105.44	568.62	189	! 09vd16 drosophila melanogaste
sp_organelle:Q9MF32	-	7.00	105.44	568.62	189	! 09mf32 brachyderes griseocens.
sp_organelle:Q9MF30	-	7.00	105.44	568.62	189	! 09mf30 brachyderes pubescens. c
sp_mammal:O02834	+	7.00	105.40	568.07	190	! 02834 sus scrofa (pig). adipoc
sp_organelle:Q9MF29	+	7.00	105.40	568.07	190	! 09mf29 brachyderes rugatus scu
sp_organelle:Q9MF28	+	7.00	105.40	568.07	190	! 09mf28 brachyderes rugatus scu
sp_organelle:Q9MF27	-	7.00	105.40	568.07	190	! 09mf27 brachyderes rugatus scu
sp_organelle:Q9MF26	-	7.00	105.40	568.07	190	! 09mf26 brachyderes rugatus scu
sp_organelle:Q9MF25	-	7.00	105.40	568.07	190	! 09mf25 brachyderes rugatus scu
sp_organelle:Q9MF24	-	7.00	105.40	568.07	190	! 09mf24 brachyderes rugatus scu
sp_organelle:Q9MF23	-	7.00	105.40	568.07	190	! 09mf23 brachyderes rugatus scu
sp_organelle:Q9MF22	-	7.00	105.40	568.07	190	! 09mf22 brachyderes rugatus scu
sp_organelle:Q9MF21	-	7.00	105.40	568.07	190	! 09mf21 brachyderes rugatus scu
sp_organelle:Q9MF20	-	7.00	105.40	568.07	190	! 09mf20 brachyderes rugatus scu
sp_organelle:Q9MF19	-	7.00	105.40	568.07	190	! 09mf19 brachyderes rugatus scu
sp_organelle:Q9MF18	-	7.00	105.40	568.07	190	! 09mf18 brachyderes rugatus scu
sp_organelle:Q9MF17	-	7.00	105.40	568.07	190	! 09mf17 brachyderes rugatus scu
sp_organelle:Q9MF16	-	7.00	105.40	568.07	190	! 09mf16 brachyderes rugatus scu
sp_organelle:Q9MF15	-	7.00	105.40	568.07	190	! 09mf15 brachyderes rugatus ruga
sp_organelle:Q9MF14	-	7.00	105.40	568.07	190	! 09mf14 brachyderes rugatus ruga
sp_organelle:Q9MF13	-	7.00	105.40	568.07	190	! 09mf13 brachyderes rugatus ruga
sp_organelle:Q9MF12	-	7.00	105.40	568.07	190	! 09mf12 brachyderes rugatus ruga
sp_organelle:Q9MF11	-	7.00	105.40	568.07	190	! 09mf11 brachyderes rugatus ruga

sp_organelle:Q9MF10	-	7.00	105.40	568.07	190	! 09mf10 brachyderes rugatus r
sp_organelle:Q9MF09	-	7.00	105.40	568.07	190	! 09mf09 brachyderes rugatus r
sp_organelle:Q9MF08	-	7.00	105.40	568.07	190	! 09mf08 brachyderes rugatus r
sp_organelle:Q9MF07	-	7.00	105.40	568.07	190	! 09mf07 brachyderes rugatus r
sp_organelle:Q9MF06	-	7.00	105.40	568.07	190	! 09mf06 brachyderes rugatus r
sp_organelle:Q9MF05	-	7.00	105.40	568.07	190	! 09mf05 brachyderes rugatus r
sp_organelle:Q9MF04	-	7.00	105.40	568.07	190	! 09mf04 brachyderes rugatus r
sp_organelle:Q9MF03	-	7.00	105.40	568.07	190	! 09mf03 brachyderes rugatus r
sp_organelle:Q9MF02	-	7.00	105.40	568.07	190	! 09mf02 brachyderes rugatus r
sp_organelle:Q9MF01	-	7.00	105.40	568.07	190	! 09mf01 brachyderes rugatus h
sp_organelle:Q9MEZ9	-	7.00	105.40	568.07	190	! 09mez9 brachyderes rugatus h
sp_organelle:Q9MEZ8	-	7.00	105.40	568.07	190	! 09mez8 brachyderes rugatus h
sp_organelle:Q9MEZ7	-	7.00	105.40	568.07	190	! 09mez7 brachyderes rugatus h
sp_organelle:Q9MDU5	-	7.00	105.40	568.07	190	! 09md5 brachyderes rugatus h
sp_organelle:Q9MDT3	-	7.00	105.40	568.07	190	! 09mdt3 brachyderes rugatus h
sp_invertebrate:Q9VGN9	+	7.00	105.37	567.52	191	! 09vgn9 drosophila melanoga
sp_archaea:Q9VB35	+	7.00	105.34	566.97	192	! 09yb35 aeropyrum pernix. hyp
sp_vertibrate:Q919F8	-	7.00	105.31	566.43	193	! 0919f8 drosophila adelaidae.
sp_vertibrate:Q919F7	-	7.00	105.31	566.43	193	! 0919f7 basileuterus culicifo
sp_archaea:Q92961	+	7.00	105.27	565.89	194	! 029261 archaeoglobus fulgidu
sp_archaea:Q58972	+	7.00	105.27	565.89	194	! 058972 pyrococcus horikoshii
sp_bacteria:Q45576	+	7.00	105.27	565.89	194	! 045576 bacillus stearothermo
sp_bacteria:Q9RS76	+	7.00	105.27	565.89	194	! 09rs76 deinococcus radiodura
sp_plant:Q9ZPT2	+	7.00	105.21	564.83	196	! 09zpt2 arabidopsis thaliana
sp_bacteria:P95037	+	7.00	105.14	563.77	198	! 095037 mycobacterium tubercu
sp_human:Q75983	+	7.00	105.14	563.77	198	! 075983 homo sapiens (human).
sp_bacteria:Q9KSL6	-	7.00	105.08	562.73	200	! 09ksl6 vibrio cholerae. alph
sp_plant:Q9LME5	-	7.00	105.08	562.73	200	! 09lme5 oryza saliva (rice).
sp_bacteria:Q9JXE1	-	7.00	105.02	561.70	202	! 09jxe1 neisseria meningitidi
sp_bacteria:Q9JWJ9	+	7.00	105.02	561.70	202	! 09jwj9 neisseria meningitidi
sp_archaea:Q52008	+	7.00	104.95	560.68	204	! 052008 halobacterium sp. orf
sp_bacteria:Q9L3G5	+	7.00	104.95	560.68	204	! 09l3g5 klebsiella pneumoniae
sp_bacteria:Q9L6B5	+	7.00	104.92	560.18	205	! 09l6b5 versinia pestis. dime
sp_bacteria:Q9Z575	-	7.00	104.83	558.69	208	! 09z575 streptomyces coelicol
sp_human:O14811	+	7.00	104.71	556.73	212	! 014811 homo sapiens (human)
sp_virus:Q69361	+	7.00	104.68	556.25	213	! 069361 feline herpesvirus (f
sp_vertibrate:Q9PUP6	-	7.00	104.59	554.82	216	! 09pup6 colius striatus. oocy
sp_rudent:Q9H116	+	7.00	104.56	554.35	217	! 09h116 mus musculus (mouse).
sp_rudent:Q9J152	+	7.00	104.56	554.35	217	! 09j152 rattus norvegicus (ra
sp_bacteria:Q9Z4A0	+	7.00	104.47	552.95	220	! 09z4a0 streptomyces violaceo
sp_bacteria:Q9K3K8	+	7.00	104.41	552.03	222	! 09k3k8 streptomyces coelicol
sp_bacteria:Q9JQ09	+	7.00	104.41	552.03	222	! 09jq09 neisseria meningitidi
sp_bacteria:Q86695	+	7.00	104.36	551.12	224	! 086695 streptomyces coelicol
sp_rudent:Q9JF51	+	7.00	104.36	551.12	224	! 09jf51 rattus norvegicus (ra
sp_plant:Q9SH12	-	7.00	104.30	550.22	226	! 09sh12 arabidopsis thaliana
sp_bacteria:P96689	+	7.00	104.22	548.89	229	! 096689 bacillus subtilis. yd
sp_bacteria:Q84936	+	7.00	104.22	548.89	229	! 084936 streptococcus pneumoni
sp_bacteria:Q9S609	+	7.00	104.22	548.89	229	! 09s609 streptococcus pneumoni
sp_bacteria:Q9JQL5	+	7.00	104.19	548.45	230	! 09jql5 neisseria meningitidi
sp_invertebrate:Q9NNR2	-	7.00	104.08	546.71	234	! 09nnr2 leishmania major. pi
sp_human:Q9N528	+	7.00	104.05	546.28	235	! 09n528 homo sapiens (human).
sp_archaea:Q26387	+	7.00	104.00	545.43	237	! 026387 methanobacterium ther
sp_bacteria:Q9RS96	+	7.00	104.00	545.43	237	! 09rs96 deinococcus radiodura
sp_plant:Q23959	+	7.00	103.95	544.59	239	! 023959 glycine max (soybean)
sp_plant:Q9XH16	+	7.00	103.92	544.17	240	! 09xh16 oryza sativa (rice).
sp_plant:Q9XK6	+	7.00	103.89	543.75	241	! 09xk6 sorghum bicolor (sorg
sp_archaea:Q9DZ4	-	7.00	103.84	542.93	243	! 09dz4 aeropyrum pernix. 243
sp_vertibrate:Q919U0	-	7.00	103.84	542.93	243	! 0919u0 lagopus leucurus (whi
sp_vertibrate:Q919T9	-	7.00	103.84	542.93	243	! 0919t9 lagopus leucurus (whi
sp_human:Q75062	+	7.00	103.79	542.11	245	! 075062 homo sapiens (human).
sp_plant:Q9LH11	+	7.00	103.79	542.11	245	! 09lh11 arabidopsis thaliana
sp_invertebrate:Q22895	-	7.00	103.76	541.70	246	! 022895 caenorhabditis eleg
sp_bacteria:P95555	+	7.00	103.74	541.30	247	! 095555 pseudomonas stutzeri
sp_bacteria:Q9RD03	-	7.00	103.74	541.30	247	! 09rd03 streptomyces coelicol
sp_plant:Q9SVC5	+	7.00	103.74	541.30	247	! 09svc5 arabidopsis thaliana
sp_bacteria:Q9PRR7	-	7.00	103.71	540.90	248	! 09prb7 ureaplasma urealyticu
sp_bacteria:Q9PK7	-	7.00	103.71	540.90	248	! 09pk7 chlamydia muridarum.
sp_invertebrate:Q9VKG5	-	7.00	103.71	540.90	248	! 09vkg5 drosophila melanoga
sp_bacteria:Q9RXJ6	-	7.00	103.68	540.50	249	! 09rxj6 deinococcus radiodura
sp_organelle:Q9T384	+	7.00	103.68	540.50	249	! 09t384 porphyra purpurea. su
sp_bacteria:Q9RSPI	-	7.00	103.66	540.10	250	! 09rsp1 deinococcus radiodura

sp_vertibrate:Q9IGX0	-	7.00	103.63	539.70	251	Q919x0	colinus virginianus (bold)	-	7.00	101.91	513.49	329	Q9np22	homo sapiens (human)
sp_bacteria:Q9ZG68	-	7.00	103.61	539.31	252	Q92998	heliobacillus mobilis. c	-	7.00	101.83	512.35	333	Q31568	bacillus subtilis. y
sp_bacteria:P94179	-	7.00	103.53	538.13	255	P94179	anabaena sp. (strain pcd	-	7.00	101.83	512.35	333	Q9S849	arabidopsis thaliana
sp_bacteria:Q9XDX8	+	7.00	103.53	538.13	256	Q9Xdx8	rhodospirillum palustris	+	7.00	101.82	511.07	334	Q9R0d8	mus musculus (mouse)
sp_phase:Q38150	-	7.00	103.51	537.75	256	Q38150	bacteriophage sppl. phag	-	7.00	101.80	511.78	335	Q51587	plectonema boyanum.
sp_invertebrate:Q9N762	-	7.00	103.48	537.36	257	Q9n762	leishmania major. poss	-	7.00	101.80	511.78	335	Q9ks12	vibrio cholerae. vit
sp_bacteria:Q9WX18	-	7.00	103.46	536.98	258	Q9wx18	streptomyces coelicolor.	-	7.00	101.76	511.22	337	Q9Z9h3	streptomyces venezue
sp_virus:Q68105	+	7.00	103.43	536.59	259	Q68105	hawaii calicivirus. hypod	+	7.00	101.76	511.22	337	Q9Z9h3	streptomyces venezue
sp_bacteria:P96327	+	7.00	103.41	536.21	260	P96327	escherichia coli. putat	+	7.00	101.76	511.22	337	Q9Z9h3	streptomyces venezue
sp_bacteria:Q52602	+	7.00	103.36	535.46	262	Q52602	agrobacterium tumefaciens	+	7.00	101.76	511.22	337	Q9Z9h3	streptomyces venezue
sp_bacteria:Q9RJ31	-	7.00	103.36	535.46	262	Q9rj31	streptomyces coelicolor.	-	7.00	101.76	511.22	337	Q9Z9h3	streptomyces venezue
sp_invertebrate:Q18146	-	7.00	103.36	535.46	262	Q18146	caenorhabditis elegans	-	7.00	101.72	510.67	339	Q45162	caenorhabditis eleg
sp_bacteria:Q05919	-	7.00	103.34	535.08	263	Q05919	mycobacterium tuberculosis	-	7.00	101.68	510.12	341	P96358	mycobacterium tuberc
sp_human:Q9G619	-	7.00	103.34	535.08	263	Q9G619	mycobacterium tuberculosis	-	7.00	101.68	510.12	341	P96358	mycobacterium tuberc
sp_archaea:Q9F915	-	7.00	103.26	533.97	266	Q9F915	aeropyrum pernix. hypoth	-	7.00	101.65	509.57	343	Q9Xb17	bacillus cereus. 5-m
sp_plant:Q9XJ34	-	7.00	103.26	533.97	266	Q9Xj34	arabidopsis thaliana (mc	-	7.00	101.61	509.02	345	P70913	bacillus intermedius
sp_bacteria:Q9PIB8	+	7.00	103.25	532.14	271	Q9piB8	campylobacter jejuni. 2-	+	7.00	101.59	508.75	346	Q9K1e2	streptomyces hygrosc
sp_invertebrate:Q01644	-	7.00	103.15	532.14	271	Q01644	drosophila grimshawi (c	-	7.00	101.57	508.48	347	Q9Kku4	streptomyces coelic.
sp_plant:Q82115	-	7.00	103.15	532.14	271	Q82115	oryza sativa (rice). zinf	-	7.00	101.55	508.21	348	Q30270	archaeoglobus fulgid
sp_bacteria:Q9K459	-	7.00	103.03	530.35	276	Q9K459	pseudomonas putida. repl	-	7.00	101.50	507.41	351	Q52013	pseudomonas putida. l
sp_human:Q9N2P3	-	7.00	102.92	528.61	281	Q9np23	homo sapiens (human). ad	-	7.00	101.48	507.14	352	Q9Kuf6	vibrio cholerae. pro
sp_plant:Q9ZSE3	-	7.00	102.92	528.61	281	Q9Zse3	euglena gracilis. fibril	-	7.00	101.46	506.88	353	P95274	mycobacterium tuberc
sp_plant:Q9MAV9	-	7.00	102.89	528.26	282	Q9m4V9	mesembryanthemum crystal	-	7.00	101.39	505.83	357	Q17682	caenorhabditis eleg
sp_bacteria:Q9RK77	-	7.00	102.83	527.23	285	Q9rk77	streptomyces coelicolor	-	7.00	101.37	505.57	358	Q9lia0	arabidopsis thaliana
sp_invertebrate:Q22623	+	7.00	102.83	527.23	285	Q22623	caenorhabditis elegans	+	7.00	101.36	505.31	359	Q53867	mycobacterium tuberc
sp_virus:R89001	+	7.00	102.83	527.23	285	P89001	mastomys natalensis papil	+	7.00	101.36	505.31	359	Q53867	mycobacterium tuberc
sp_virus:P87595	-	7.00	102.80	526.89	286	P87595	cowpox virus (cpv). 41k	-	7.00	101.34	505.31	360	Q9X613	chlorobium tepidum. (
sp_bacteria:Q9KSW1	-	7.00	102.76	526.22	288	Q9Ksw1	bacillus halodurans. sta	-	7.00	101.34	505.31	360	Q9X613	chlorobium tepidum. (
sp_rhodent:Q9JLU8	-	7.00	102.74	525.88	289	Q9Jlu8	cavia porcellus (guinea	-	7.00	101.34	505.31	360	Q9X613	chlorobium tepidum. (
sp_bacteria:Q9JU76	+	7.00	102.69	525.22	291	Q9Ju76	neisseria meningitidis (c	-	7.00	101.30	504.54	362	Q27744	plasmidium falcipar
sp_human:Q9NP24	-	7.00	102.69	525.22	291	Q9np24	homo sapiens (human). ad	-	7.00	101.30	504.54	362	Q27744	plasmidium falcipar
sp_plant:Q9X533	-	7.00	102.58	523.57	296	Q9x533	zea mays (maize). hypoth	-	7.00	101.30	504.54	362	Q9U6r2	plasmidium falcipar
sp_invertebrate:Q18574	-	7.00	102.56	523.25	297	Q18574	caenorhabditis elegans	-	7.00	101.25	503.77	365	Q9R9q2	rhizobium meliloti (s
sp_bacteria:Q9K9H3	-	7.00	102.54	522.92	298	Q9K9h3	bacillus halodurans. 6-b	-	7.00	101.23	503.52	366	Q9Kxn5	streptomyces coelicol
sp_human:Q04391	-	7.00	102.54	522.92	298	Q04391	homo sapiens (human). sy	-	7.00	101.22	503.27	367	Q9Kps3	vibrio cholerae. thi
sp_human:Q00173	-	7.00	102.54	522.92	298	Q00173	homo sapiens (human). sy	-	7.00	101.22	503.27	367	Q9Kps3	vibrio cholerae. thi
sp_human:Q00560	+	7.00	102.54	522.92	298	Q00560	homo sapiens (human). me	-	7.00	101.22	503.27	367	Q9S632	oryza sativa (rice).
sp_virus:Q9J4P3	+	7.00	102.54	522.92	298	Q9J4P3	human immunodeficiency v	+	7.00	101.18	502.76	369	Q12593	cryphonectria parasit
sp_bacteria:Q9KK85	+	7.00	102.52	522.60	299	Q9K85	brevibacterium linens. f	+	7.00	101.18	502.76	369	Q12593	cryphonectria parasit
sp_bacteria:Q69914	+	7.00	102.46	521.64	302	Q69914	streptomyces coelicolor.	+	7.00	101.16	502.51	370	Q29775	archaeoglobus fulgid
sp_bacteria:Q9X7M7	+	7.00	102.46	521.64	302	Q9X7M7	streptomyces coelicolor.	+	7.00	101.16	502.51	370	Q9Rtu4	deinococcus radiodur
sp_bacteria:Q9LOJ0	+	7.00	102.46	521.64	302	Q9LOJ0	streptomyces coelicolor.	+	7.00	101.16	502.51	370	Q9Rtu4	deinococcus radiodur
sp_mammal:Q9N0U9	+	7.00	102.46	521.64	302	Q9n0U9	ovis aries (sheep). amil	+	7.00	101.13	502.01	372	Q33109	saccharaterium leprae
sp_archaea:Q9V9D5	+	7.00	102.44	521.33	303	Q9V9d5	aeropyrum pernix. hypoth	+	7.00	101.13	502.01	372	Q33109	saccharaterium leprae
sp_bacteria:P71913	-	7.00	102.41	521.01	304	P71913	mycobacterium tuberculosis	-	7.00	101.13	502.01	372	Q33109	saccharaterium leprae
sp_archaea:Q9V9J0	-	7.00	102.39	520.69	305	Q9V9J0	aeropyrum pernix. 305aa	-	7.00	101.11	501.77	373	Q58859	methanococcus jannasc
sp_bacteria:Q53929	-	7.00	102.37	520.38	306	Q53929	streptomyces coelicolor.	-	7.00	101.11	501.77	373	Q53253	mycobacterium tubercu
sp_archaea:Q26926	+	7.00	102.35	520.07	307	Q26926	methanobacterium thermo	+	7.00	101.06	501.03	376	Q85136	rhodobacter sphaeroi
sp_archaea:Q9VDY2	+	7.00	102.33	519.76	308	Q9Vdy2	aeropyrum pernix. hypoth	+	7.00	101.03	500.54	378	Q01451	haemophilus influenza
sp_plant:Q82583	-	7.00	102.31	519.45	309	Q82583	zantedeschia aethiopica	-	7.00	101.01	500.30	379	Q9T9q1	caenorhabditis eleg
sp_archaea:Q9UXJ4	-	7.00	102.29	519.14	310	Q9uxJ4	sulfolobus solfataricus.	+	7.00	100.99	500.05	380	Q43568	nicotiana tabacum (co
sp_bacteria:Q67437	+	7.00	102.29	519.14	310	Q67437	aquifex aeolicus. rna po	+	7.00	100.96	499.57	382	Q69815	streptomyces coelicol
sp_bacteria:Q33200	+	7.00	102.27	518.83	311	Q33200	mycobacterium tuberculosis	+	7.00	100.91	498.85	382	Q9Rxp9	deinococcus radiodura
sp_bacteria:Q9R0H1	-	7.00	102.27	518.83	311	Q9rdh1	streptomyces coelicolor.	-	7.00	100.91	498.85	385	Q9X932	streptomyces coelicol
sp_bacteria:Q9RX37	-	7.00	102.27	518.83	311	Q9rx37	deinococcus radiodurans.	-	7.00	100.91	498.85	385	Q9X932	streptomyces coelicol
sp_invertebrate:Q77442	+	7.00	102.25	518.53	312	Q77442	taenia saginata. heat	+	7.00	100.81	497.44	391	Q44032	alcaligenes eutrophus
sp_bacteria:Q9KV12	+	7.00	102.21	517.92	314	Q9Kv12	vibrio cholerae. trna de	+	7.00	100.80	497.20	392	Q9S1a1	arabidopsis thaliana
sp_plant:Q9M8F9	+	7.00	102.19	517.61	315	Q9m8F9	arabidopsis thaliana (mc	+	7.00	100.76	496.74	394	Q9X716	pseudomonas sp. p-hyd
sp_bacteria:Q9X832	+	7.00	102.09	516.12	320	Q9X832	streptomyces coelicolor.	+	7.00	100.76	496.74	394	Q9X716	pseudomonas sp. p-hyd
sp_plant:Q9SJV0	-	7.00	102.09	516.12	320	Q9sjV0	arabidopsis thaliana (mc	-	7.00	100.76	496.74	394	Q9X716	pseudomonas sp. p-hyd
sp_bacteria:Q69128	-	7.00	102.07	515.82	321	Q69128	lythrodia pseudomalle	-	7.00	100.75	496.50	395	Q14828	homo sapiens (human)
sp_bacteria:Q9JPC8	-	7.00	102.07	515.82	321	Q9Jpc8	rhodocyclus gelatinosus	-	7.00	100.75	496.50	395	Q14828	homo sapiens (human)
sp_human:Q99608	-	7.00	102.07	515.82	321	Q9Jpc8	rhodocyclus gelatinosus	-	7.00	100.75	496.50	395	Q14828	homo sapiens (human)
sp_plant:Q9SE17	+	7.00	102.07	515.82	321	Q9sel7	arabidopsis thaliana. ne	+	7.00	100.75	496.50	395	Q9R615	agrobacterium tumefac
sp_bacteria:Q9KWL1	+	7.00	102.03	515.23	323	Q9Kw1	vibrio cholerae. d-3-pho	+	7.00	100.72	496.04	397	Q53870	mycobacterium tubercu
sp_human:Q9P0V3	+	7.00	102.03	515.23	323	Q9P0V3	homo sapiens (human). se	+	7.00	100.72	496.04	397	Q53870	mycobacterium tubercu
sp_bacteria:Q32412	-	7.00	102.01	514.94	324	Q32412	rhodospirillum molischia	-	7.00	100.72	496.04	397	Q53870	mycobacterium tubercu
sp_bacteria:Q9JQV2	-	7.00	102.01	514.94	324	Q9Jqv2	neisseria meningitidis (c	+	7.00	100.70	495.81	398	Q919x1	colinus virginianus (
sp_human:Q15630	-	7.00	102.01	514.94	324	Q15630	homo sapiens (human). tr	-	7.00	100.67	495.36	400	Q9q779	rattus norvegicus (ra
sp_bacteria:Q9X892	+	7.00	101.95	514.07	327	Q9X892	streptomyces coelicolor.	+	7.00	100.65	495.13	401	Q9ek9	caenorhabditis eleg
sp_invertebrate:Q20759	+	7.00	101.93	513.78	328	Q20759	caenorhabditis elegans	+	7.00	100.64	494.90	402	Q23054	arabidopsis thaliana
sp_bacteria:P96645	+	7.00	101.91	513.49	329	P96645	bacillus subtilis. yddh	+	7.00	100.62	494.68	403	Q42417	gallus gallus (chicke

sp_vertibrate:Q919K9	+	7.00	100.62	494.68	403	!	Q919k9 gallus gallus (chicken).	sp_bacteria:O66573	-	7.00	99.56	479.75	476	!	O66573 aquilex aeolicus. ald
sp_rodent:Q9JIK3	+	7.00	100.57	494.00	406	!	Q9jik3 rattus norvegicus (rat).	sp_bacteria:O83837	-	7.00	99.52	479.20	479	!	O83837 treponema pallidum. h
sp_plant:Q9M2M8	+	7.00	100.53	493.33	409	!	Q9m2m8 arabidopsis thaliana (md	sp_human:O92743	+	7.00	99.51	479.02	480	!	O92743 homo sapiens (human).
sp_plant:Q9M9Z1	-	7.00	100.48	492.67	412	!	Q9m9z1 arabidopsis thaliana (md	sp_rodent:Q9R118	+	7.00	99.51	479.02	480	!	Q9r118 mus musculus (mouse).
sp_bacteria:Q9RYF0	+	7.00	100.43	492.01	415	!	Q9ryf0 deinococcus radiodurans.	sp_rodent:Q9QZK6	+	7.00	99.51	479.02	480	!	Q9qzk6 mus musculus (mouse).
sp_plant:Q9SJH8	+	7.00	100.43	492.01	415	!	Q9sjh8 arabidopsis thaliana (md	sp_rodent:Q9QZK5	+	7.00	99.51	479.02	480	!	Q9qzk5 rattus norvegicus (ra
sp_virus:Q96626	+	7.00	100.43	492.01	415	!	Q96626 equine herpesvirus 2. o	sp_bacteria:O66144	+	7.00	99.50	478.83	481	!	O66144 citrobacter amalonati
sp_invertebrate:Q17495	+	7.00	100.40	491.58	417	!	Q17495 caenorhabditis elegans	sp_bacteria:Q9RWF2	+	7.00	99.48	478.65	482	!	O9rwf2 deinococcus radiodura
sp_archaea:Q9UY07	+	7.00	100.37	491.15	419	!	Q9uy07 pyrococcus abyssi. gluc	sp_bacteria:O51131	+	7.00	99.47	478.47	483	!	O51131 borrelia burgdorferi
sp_invertebrate:Q24548	+	7.00	100.37	491.15	419	!	Q24548 drosophila melanogaste	sp_invertebrate:Q91239	+	7.00	99.44	478.10	485	!	P91239 caenorhabditis eleg
sp_plant:Q91JUT4	+	7.00	100.37	491.15	419	!	Q9lut4 arabidopsis thaliana (md	sp_invertebrate:Q9V9N1	-	7.00	99.44	478.10	485	!	Q9v9n1 drosophila melanoga
sp_human:Q9NP21	+	7.00	100.34	490.72	421	!	Q9np21 homo sapiens (human).	sp_bacteria:Q9RFF6	+	7.00	99.43	477.92	486	!	Q9rf66 rhodobacter sphaeroi
sp_bacteria:O51572	+	7.00	100.31	490.29	423	!	Q51572 borrelia burgdorferi (ly	sp_bacteria:O50436	+	7.00	99.40	477.56	488	!	O50436 mycobacterium tubercu
sp_human:Q9POY4	+	7.00	100.31	490.29	423	!	Q9poy4 homo sapiens (human).	sp_mammal:Q28425	+	7.00	99.31	476.31	495	!	Q28425 gorilla gorilla gorill
sp_plant:Q9SM16	+	7.00	100.30	490.07	424	!	Q9sm16 zea mays (maize).	sp_plant:Q9SCN2	+	7.00	99.27	475.78	498	!	Q9scn2 arabidopsis thaliana
sp_bacteria:Q9N336	+	7.00	100.28	489.86	425	!	Q9zn36 helicobacter pylori j99	sp_invertebrate:O76323	+	7.00	99.21	474.91	503	!	O76323 loligo pealeii (lon
sp_bacteria:Q9RYM5	+	7.00	100.24	489.23	428	!	Q9rym5 deinococcus radiodurans.	sp_virus:O82010	+	7.00	99.21	474.91	503	!	O82010 human papillomavirus
sp_invertebrate:Q9VYA9	-	7.00	100.24	489.23	428	!	Q9vyay9 drosophila melanogaste	sp_bacteria:Q9ZC68	+	7.00	99.19	474.56	505	!	Q9zc68 yersinia pestis. arud
sp_bacteria:Q9PBC1	+	7.00	100.21	488.81	430	!	Q9pbc1 xylella fastidiosa. thre	sp_rodent:O62741	+	7.00	99.17	474.39	506	!	O62741 rattus norvegicus (ra
sp_bacteria:Q9RYX0	+	7.00	100.21	488.81	430	!	Q9kys0 streptomyces coelicolor.	sp_bacteria:O66973	+	7.00	99.12	473.70	510	!	O66973 aquilex aeolicus. l-a
sp_rodent:O54844	-	7.00	100.15	487.98	434	!	O54844 mus musculus (mouse).	sp_virus:Q91DV8	+	7.00	99.11	473.53	511	!	Q91dv8 human immunodeficien
sp_fungi:Q9P452	-	7.00	100.13	487.77	435	!	Q9p452 sclerotinia sclerotiorum	sp_invertebrate:O9U023	-	7.00	99.10	473.36	512	!	O9u023 giardia lamblia (gi
sp_plant:Q92778	+	7.00	100.13	487.77	435	!	Q9sm17 zea mays (maize).	sp_bacteria:Q9KLSA	+	7.00	99.09	473.19	513	!	Q9kls4 vibrio cholerae. sodi
sp_plant:Q92718	+	7.00	100.12	487.56	436	!	Q92778 arabidopsis thaliana (md	sp_fungi:Q9Y783	+	7.00	99.09	473.19	513	!	Q9y7p3 schizosaccharomyces p
sp_plant:O82244	+	7.00	100.09	487.15	438	!	Q82244 arabidopsis thaliana (md	sp_bacteria:Q9RBP2	+	7.00	99.07	473.02	514	!	Q9rbp2 rhodococcus sp. ad45.
sp_virus:Q84402	+	7.00	100.08	486.95	439	!	Q84402 peanut bud necrosis viru	sp_plant:Q22509	+	7.00	99.05	472.69	516	!	Q22509 arabidopsis thaliana
sp_bacteria:Q9X8S1	+	7.00	100.05	486.54	441	!	Q9x8s1 streptomyces coelicolor.	sp_plant:Q9SYC2	+	7.00	99.05	472.69	516	!	Q9sydc arabidopsis thaliana
sp_human:Q9NP19	+	7.00	99.98	485.53	446	!	Q9np19 homo sapiens (human).	sp_plant:Q9ZU50	+	7.00	99.04	472.52	517	!	Q9zu50 arabidopsis thaliana
sp_plant:Q91LVA0	+	7.00	99.98	485.53	446	!	Q91lva0 arabidopsis thaliana (md	sp_archaea:O27324	+	7.00	99.01	472.18	519	!	O27324 methanobacterium ther
sp_bacteria:Q9A824	+	7.00	99.96	485.33	447	!	P74824 sphingomonas sp. s88. hy	sp_bacteria:Q9P2042	+	7.00	99.01	472.18	519	!	P72042 mycobacterium tubercu
sp_bacteria:Q9XKZ6	+	7.00	99.96	485.33	447	!	Q9xkz6 streptomyces coelicolor.	sp_bacteria:Q9PCG3	+	7.00	99.01	472.18	519	!	Q9pcg3 xylella fastidiosa. 2
sp_bacteria:Q9XKX8	+	7.00	99.93	484.94	449	!	Q9xkx8 escherichia coli O157:h7	sp_virus:Q9YQ22	+	7.00	99.01	472.18	519	!	Q9yq22 bovine calicivirus. 2
sp_human:Q9X934	+	7.00	99.93	484.94	449	!	O43934 homo sapiens (human).	sp_plant:O48881	-	7.00	98.96	471.51	523	!	O48881 brassica napus (rape)
sp_bacteria:Q9N15	+	7.00	99.93	484.94	449	!	Q9n15 homo sapiens (human).	sp_plant:O9LV45	-	7.00	98.91	470.85	527	!	Q9lv45 arabidopsis thaliana
sp_bacteria:Q92596	+	7.00	99.92	484.74	450	!	Q92596 helicobacter pylori (cam	sp_bacteria:O69547	-	7.00	98.88	470.36	530	!	O69547 mycobacterium leprae.
sp_bacteria:Q91554	+	7.00	99.92	484.74	450	!	Q91554 streptomyces galileus.	sp_bacteria:Q9R9Y7	+	7.00	98.88	470.36	530	!	Q9ry7 pseudomonas putida. t
sp_human:Q9NP20	-	7.00	99.90	484.54	451	!	Q9np20 homo sapiens (human).	sp_plant:Q9SK21	+	7.00	98.87	470.20	531	!	Q9sk21 arabidopsis thaliana
sp_plant:Q9NM16	+	7.00	99.90	484.54	451	!	Q9nm16 arabidopsis thaliana (md	sp_bacteria:Q49972	+	7.00	98.84	469.87	533	!	O49972 mycobacterium leprae.
sp_plant:Q9W9D7	+	7.00	99.89	484.34	452	!	Q9w9d7 arabidopsis thaliana (md	sp_fungi:O74936	+	7.00	98.83	469.71	534	!	O74936 yarrowia lipolytica.
sp_bacteria:Q968903	+	7.00	99.85	483.75	455	!	Q968903 actinonema pretiosum	sp_invertebrate:Q9W017	+	7.00	98.82	469.55	535	!	O9w017 drosophila melanoga
sp_human:Q9UP88	+	7.00	99.85	483.75	455	!	Q9up88 homo sapiens (human).	sp_virus:O56650	+	7.00	98.81	469.39	536	!	O56650 adeno-associated virus
sp_bacteria:Q9KUF5	-	7.00	99.83	483.56	456	!	Q9kuf5 vibrio cholerae. proteas	sp_invertebrate:Q9VPM5	+	7.00	98.77	468.91	539	!	Q9vpm5 drosophila melanoga
sp_human:Q9Y215	-	7.00	99.83	483.56	456	!	Q9y215 homo sapiens (human).	sp_virus:Q66287	+	7.00	98.77	468.91	539	!	Q66287 unidentified. capsid
sp_invertebrate:Q20084	-	7.00	99.83	483.56	456	!	Q20084 caenorhabditis elegans	sp_virus:Q9YQ50	+	7.00	98.77	468.91	539	!	Q9yq50 human calicivirus. ca
sp_archaea:Q9YE02	+	7.00	99.81	483.17	458	!	Q9yeu2 aeropyrum pernix. 458a	sp_virus:Q68537	+	7.00	98.77	468.91	539	!	O68537 human calicivirus. an
sp_bacteria:Q966229	+	7.00	99.81	483.17	458	!	Q966229 escherichia coli. 458a	sp_virus:Q9WR21	+	7.00	98.77	468.91	539	!	Q9wr21 norwalk-like virus sp
sp_bacteria:Q9RMU4	+	7.00	99.81	483.17	458	!	Q9rmu4 klebsiella pneumoniae. n	sp_virus:Q9WR20	+	7.00	98.77	468.91	539	!	Q9wr20 norwalk-like virus sp
sp_human:O43464	+	7.00	99.81	483.17	458	!	O43464 homo sapiens (human).	sp_virus:Q9WR29	+	7.00	98.77	468.91	539	!	Q9wry9 norwalk-like virus sp
sp_rodent:Q9R108	+	7.00	99.81	483.17	458	!	Q9r108 mus musculus (mouse).	sp_virus:Q9WR18	+	7.00	98.77	468.91	539	!	Q9wry8 norwalk-like virus sp
sp_rodent:Q9J1V5	+	7.00	99.81	483.17	458	!	Q9j1v5 mus musculus (mouse).	sp_virus:Q9W9Y9	+	7.00	98.77	468.91	539	!	Q9w9y9 norwalk-like virus sp
sp_bacteria:Q9ZCF1	+	7.00	99.79	482.98	459	!	Q9zcf1 rickettsia prowazekii. d	sp_virus:Q9W9P8	+	7.00	98.77	468.91	539	!	Q9w9t8 norwalk-like virus sp
sp_bacteria:Q9X8I8	+	7.00	99.79	482.98	459	!	Q9x8i8 streptomyces coelicolor.	sp_virus:Q9IV44	+	7.00	98.77	468.91	539	!	Q9iv44 human calicivirus hu/
sp_plant:Q9LGY8	+	7.00	99.79	482.98	459	!	Q9lgy8 oryza sativa (rice).	sp_virus:Q9IV38	+	7.00	98.77	468.91	539	!	Q9iv38 human calicivirus hu/
sp_bacteria:Q96236	+	7.00	99.78	482.78	460	!	O66236 escherichia coli. phosph	sp_bacteria:O05692	+	7.00	98.76	468.75	540	!	O05692 rhizobium leguminosar
sp_bacteria:Q66243	+	7.00	99.78	482.78	460	!	O66243 klebsiella pneumoniae. p	sp_human:Q60230	-	7.00	98.76	468.75	540	!	O60230 homo sapiens (human).
sp_bacteria:Q95428	+	7.00	99.78	482.78	460	!	Q95428 escherichia coli. and k1	sp_human:Q14033	-	7.00	98.74	468.43	542	!	Q14033 homo sapiens (human).
sp_bacteria:Q9XKX4	-	7.00	99.78	482.78	460	!	Q9xkx4 chryseomys picta. nadh de	sp_bacteria:O51371	-	7.00	98.72	468.27	543	!	O51371 pseudomonas aeruginos
sp_organellie:Q9XPH7	+	7.00	99.78	482.78	460	!	Q9xph7 chelonina mydas caraniga	sp_fungi:Q13449	+	7.00	98.70	467.95	545	!	O13449 saccharomyces cerevis
sp_virus:Q97990	+	7.00	99.78	482.78	460	!	Q97990 human immunodeficiency v	sp_plant:Q23944	+	7.00	98.70	467.95	545	!	O23944 fragaria vesca. cysta
sp_virus:Q96987	+	7.00	99.78	482.78	460	!	O69887 human immunodeficiency v	sp_plant:Q40451	+	7.00	98.69	467.80	546	!	Q40451 nicotiana tabacum (co
sp_bacteria:Q96954	+	7.00	99.71	481.82	465	!	O69954 streptomyces coelicolor.	sp_invertebrate:Q18041	+	7.00	98.67	467.48	548	!	Q18041 caenorhabditis eleg
sp_bacteria:Q95246	+	7.00	99.70	481.63	466	!	Q95246 rickettsia tsutsugamushi	sp_bacteria:O86291	+	7.00	98.65	467.32	549	!	O86291 mycobacterium tubercu
sp_bacteria:Q95247	+	7.00	99.70	481.63	466	!	Q95247 rickettsia tsutsugamushi	sp_bacteria:O86752	+	7.00	98.61	466.70	553	!	O86752 streptomyces coelicol
sp_bacteria:Q95249	+	7.00	99.70	481.63	466	!	Q95249 rickettsia tsutsugamushi	sp_invertebrate:Q24612	-	7.00	98.52	465.47	561	!	Q24612 drosophila pseudoot
sp_bacteria:Q95251	+	7.00	99.70	481.63	466	!	Q95251 rickettsia tsutsugamushi	sp_bacteria:Q52413	-	7.00	98.45	465.16	563	!	O52413 pseudomonas sp. alpha
sp_bacteria:Q05335	+	7.00	99.70	481.63	466	!	O05335 rickettsia typhi. 47 kda	sp_plant:Q9SV44	+	7.00	98.45	464.71	566	!	Q9sv44 arabidopsis thaliana
sp_invertebrate:Q76322	+	7.00	99.70	481.63	466	!	O76322 loligo pealeii (longfi	sp_bacteria:Q9R9H2	+	7.00	98.45	464.56	567	!	Q9r9h2 pseudomonas chlororaf
sp_bacteria:Q915J6	+	7.00	99.68	481.44	467	!	Q915j6 salmonella typhi. hypoch	sp_fungi:Q9R871	+	7.00	98.45	464.56	567	!	Q9r871 schizosaccharomyces f
sp_bacteria:Q9X803	+	7.00	99.66	481.06	469	!	Q9x803 streptomyces coelicolor.	sp_phage:Q9XJK0	-	7.00	98.44	464.41	568	!	Q9xjk0 bacteriophage 933w. a
sp_bacteria:Q96604	+	7.00	99.66	481.06	469	!	O66604 mycobacterium tuberculo	sp_archaea:Q92188	-	7.00	98.43	464.26	569	!	O92188 archaeoglobus fulgidu
sp_plant:Q9S9A3	-	7.00	99.66	481.06	469	!	Q9s9a3 lycopersicon esculentum	sp_bacteria:Q9KAT1	-	7.00	98.43	464.26	569	!	O9kat1 bacillus halodurans.
sp_bacteria:Q9K7G0	-	7.00	99.58	479.94	475	!	Q9k7g0 bacillus halodurans. lei	sp_bacteria:Q9KUF5	-	7.00	98.40	463.96	571	!	O9kuf5 bacillus halodurans.
sp_invertebrate:Q21555	-	7.00	99.58	479.94	475	!	Q21555 caenorhabditis elegans	sp_plant:Q24022	-	7.00	98.40	463.96	571	!	Q24022 lycopersicon esculent

sp_invertebrate:Q9W2P2	+	7.00	98.39	463.81	572	! Q9w2p2 drosophila melanogaster	sp_archaea:Q9UWN2	+	7.00	96.76	442.46	739	! Q9uwn2 thermococcus sp. b10
sp_invertebrate:Q9N4Q1	-	7.00	98.37	463.51	574	! Q9n4q1 caenorhabditis elegans	sp_plant:Q9S4B20	-	7.00	96.76	442.46	739	! Q9s4b20 arabidopsis thaliana
sp_bacteria:Q9K3P9	-	7.00	98.34	463.07	577	! Q9k3p9 streptomyces coelicolor	sp_plant:Q9WB58	-	7.00	96.72	441.91	744	! Q9wb58 arabidopsis thaliana
sp_bacteria:Q9X6M9	+	7.00	98.29	462.48	581	! Q9x6m9 bartonella clarridgeae	sp_plant:Q9SP17	-	7.00	96.71	441.80	745	! Q9sp17 arabidopsis thaliana
sp_invertebrate:Q23318	+	7.00	98.29	462.48	581	! Q23318 caenorhabditis elegans	sp_plant:Q9L139	-	7.00	96.71	441.80	745	! Q9l139 arabidopsis thaliana
sp_bacteria:Q9RG4	+	7.00	98.27	462.19	583	! Q9rg4 bordetella bronchiseptica	sp_bacteria:P73687	-	7.00	96.68	441.37	749	! P73687 synecocystis sp. (s)
sp_bacteria:Q9RG3	+	7.00	98.27	462.19	583	! Q9rg3 bordetella bronchiseptica	sp_bacteria:P96316	-	7.00	96.65	441.04	752	! P96316 cellivibrio gilvus. b
sp_invertebrate:Q9X466	+	7.00	98.27	462.19	583	! Q9x466 caenorhabditis elegans	sp_human:Q94993	+	7.00	96.64	440.93	753	! Q94993 homo sapiens (human)
sp_plant:P94017	+	7.00	98.26	462.04	584	! P94017 arabidopsis thaliana (mc)	sp_plant:Q9JFF1	+	7.00	96.64	440.83	754	! Q9jff1 arabidopsis thaliana
sp_plant:Q9LMD2	+	7.00	98.25	461.90	585	! Q9lmd2 arabidopsis thaliana (mc)	sp_fungi:Q74253	+	7.00	96.51	439.23	769	! Q74253 pycnoporus cinnabari
sp_bacteria:Q9KLS8	-	7.00	98.23	461.61	587	! Q9kls8 vibrio cholerae. methyl	sp_invertebrate:Q20440	+	7.00	96.51	439.23	769	! Q20440 caenorhabditis eleg
sp_invertebrate:Q00904	+	7.00	98.23	461.61	587	! Q00904 leishmania major. l417	sp_human:Q9NVB2	+	7.00	96.40	437.88	782	! Q9nvb2 homo sapiens (human)
sp_plant:Q9SVE0	+	7.00	98.23	461.61	587	! Q9sve0 arabidopsis thaliana (mc)	sp_virus:Q9WFB5	+	7.00	96.36	437.37	787	! Q9wfb5 snow goose hepatitis
sp_invertebrate:Q9YH20	-	7.00	98.23	461.61	587	! Q9yh20 gallus gallus (chicken)	sp_virus:Q9WFB2	+	7.00	96.36	437.37	787	! Q9wfb2 snow goose hepatitis
sp_bacteria:Q9XAH5	+	7.00	98.22	461.46	588	! Q9xah5 streptomyces coelicolor	sp_virus:Q9WFA8	+	7.00	96.36	437.37	787	! Q9wfa8 snow goose hepatitis
sp_invertebrate:Q02007	-	7.00	98.22	461.46	588	! Q02007 drosophila simulans (f)	sp_virus:Q9WFA5	+	7.00	96.36	437.37	787	! Q9wfa5 snow goose hepatitis
sp_bacteria:Q9ZBR4	+	7.00	98.21	461.32	589	! Q9zbr4 streptomyces coelicolor	sp_virus:Q9WFA2	+	7.00	96.36	437.37	787	! Q9wfa2 snow goose hepatitis
sp_plant:Q38801	+	7.00	98.20	461.17	590	! Q38801 arabidopsis thaliana (mc)	sp_invertebrate:Q9W27	+	7.00	96.31	436.75	793	! Q9w27 drosophila melanog
sp_rodent:Q9QWJ6	-	7.00	98.20	461.17	590	! Q9qwj6 rattus norvegicus (rat)	sp_plant:Q9LIT9	+	7.00	96.31	436.75	793	! Q9lit9 oryza sativa (rice)
sp_invertebrate:Q77267	+	7.00	98.17	460.89	592	! Q77267 drosophila melanogaster	sp_invertebrate:Q16145	+	7.00	96.29	436.45	796	! Q16145 caenorhabditis eleg
sp_archaea:Q9UYA4	+	7.00	98.14	460.46	595	! Q9uya4 pyrococcus abyssi. hypot	sp_plant:Q23190	+	7.00	96.27	436.15	799	! Q23190 arabidopsis thaliana
sp_bacteria:Q9PEF6	+	7.00	98.13	460.32	596	! Q9pef6 xylella fastidiosa. succ	sp_invertebrate:Q9VAK6	+	7.00	96.23	435.65	804	! Q9vak6 drosophila melanog
sp_rodent:Q35727	+	7.00	98.12	460.17	597	! Q35727 mus musculus (mouse). fa	sp_plant:Q9X107	+	7.00	96.20	435.35	807	! Q9x107 arabidopsis thaliana
sp_human:Q75017	+	7.00	98.11	460.03	598	! Q75017 homo sapiens (human). le	sp_invertebrate:Q44969	+	7.00	96.19	435.15	809	! Q44969 caenorhabditis eleg
sp_plant:Q9SA65	-	7.00	98.10	459.89	599	! Q9sa65 arabidopsis thaliana (mc)	sp_bacteria:Q55799	+	7.00	96.16	434.76	813	! Q55799 synecocystis sp. (s)
sp_fungi:Q9P8M8	-	7.00	98.04	459.05	605	! Q9p8m8 aspergillus niger. gluc	sp_invertebrate:P90882	-	7.00	95.96	432.34	838	! P90882 caenorhabditis eleg
sp_invertebrate:Q9TVK2	+	7.00	97.99	458.49	609	! Q9tvk2 caenorhabditis elegans	sp_invertebrate:Q10652	-	7.00	95.96	432.25	838	! Q10652 caenorhabditis eleg
sp_invertebrate:Q9VM61	+	7.00	97.97	458.22	611	! Q9vm61 drosophila melanogaster	sp_invertebrate:Q9SRR2	+	7.00	95.94	432.06	841	! Q9srr2 arabidopsis thaliana
sp_plant:Q9SPX3	+	7.00	97.94	457.80	614	! Q9spx3 arabidopsis thaliana (mc)	sp_invertebrate:Q61123	+	7.00	95.93	431.87	843	! Q61123 antiopeba histolyt
sp_vertebrate:Q9W721	+	7.00	97.92	457.53	616	! Q9w721 brachydanio rerio (zebra	sp_plant:P93068	+	7.00	95.90	431.59	846	! P93068 arabidopsis thaliana
sp_bacteria:Q9S6F5	+	7.00	97.91	457.39	617	! Q9s6f5 streptomyces coelicolor	sp_plant:P93068	+	7.00	95.90	431.49	847	! P93068 brassica oleracea (ca
sp_virus:Q56651	+	7.00	97.87	456.85	621	! Q56651 adeno-associated virus 2	sp_plant:Q04383	+	7.00	95.90	431.49	847	! Q04383 brassica oleracea (ca
sp_virus:Q89268	+	7.00	97.87	456.85	621	! Q89268 adeno-associated virus 2	sp_invertebrate:Q9U4G5	+	7.00	95.86	431.03	852	! Q9u4g5 drosophila melanog
sp_fungi:Q06451	+	7.00	97.86	456.71	622	! Q06451 saccharomyces cerevisiae	sp_fungi:Q9Y6Z8	+	7.00	95.84	430.84	854	! Q9y6z8 aspergillus oryzae. f
sp_invertebrate:Q9W4A2	-	7.00	97.85	456.58	623	! Q9w4a2 drosophila melanogaster	sp_fungi:Q9Y6Z8	+	7.00	95.84	430.75	855	! Q9y6z8 streptomyces coelicol
sp_virus:Q56136	+	7.00	97.85	456.58	623	! Q56136 adeno-associated virus 6	sp_fungi:Q002290	+	7.00	95.80	430.29	860	! Q002290 neocallimastix patri
sp_virus:Q56136	+	7.00	97.85	456.58	623	! Q56136 adeno-associated virus 6	sp_invertebrate:Q18109	+	7.00	95.70	429.01	874	! Q18109 caenorhabditis eleg
sp_plant:Q9WAP7	+	7.00	97.85	456.58	623	! Q9wap7 adeno-associated virus 1	sp_invertebrate:P91859	+	7.00	95.67	428.65	878	! P91859 caenorhabditis eleg
sp_plant:Q9ZUJ3	+	7.00	97.85	456.58	623	! Q9zuj3 arabidopsis thaliana (mc)	sp_virus:Q83101	+	7.00	95.67	428.65	878	! Q83101 unidentified leishma
sp_fungi:Q9P3M2	+	7.00	97.73	454.98	635	! Q9p3m2 neurospora crassa. conse	sp_virus:Q83101	+	7.00	95.67	428.65	878	! Q83101 unidentified leishma
sp_mammal:Q28423	+	7.00	97.73	454.98	635	! Q28423 gorilla gorilla gorilla (mc)	sp_rodent:Q9WY91	+	7.00	95.66	428.56	879	! Q9wy91 mus musculus (mouse)
sp_plant:Q9S138	+	7.00	97.73	454.98	635	! Q9s138 arabidopsis thaliana (mc)	sp_bacteria:Q9WYA9	-	7.00	95.64	428.38	881	! Q9wya9 thermotoga maritima.
sp_human:Q9NSG0	+	7.00	97.65	453.93	643	! Q9nsg0 homo sapiens (human). hy	sp_fungi:Q13981	-	7.00	95.62	428.02	885	! Q13981 schizosaccharomyces f
sp_bacteria:Q9KXB9	+	7.00	97.63	453.67	645	! Q9kxb9 escherichia coli o157:h7	sp_plant:Q9Z0X6	-	7.00	95.59	427.76	888	! Q9z0x6 arabidopsis thaliana
sp_invertebrate:Q9NE6	-	7.00	97.63	453.67	645	! Q9nee6 leishmania major. hyp	sp_bacteria:Q53121	+	7.00	95.56	427.31	893	! Q53121 mycobacterium leprae
sp_phage:Q9XJK8	+	7.00	97.63	453.67	645	! Q9xjk8 bacteriophage 933w. and	sp_human:Q9NRY3	+	7.00	95.53	426.96	897	! Q9nry3 homo sapiens (human)
sp_bacteria:Q70997	-	7.00	97.61	453.41	647	! Q70997 bacillus subtilis. hypot	sp_fungi:Q9PD43	+	7.00	95.51	426.70	900	! Q9pd43 zygosaccharomyces rou
sp_bacteria:Q9KK07	+	7.00	97.59	453.16	649	! Q9kkq7 vibrio cholerae. pts sys	sp_invertebrate:Q18260	-	7.00	95.47	426.27	905	! Q18260 caenorhabditis eleg
sp_invertebrate:Q20442	+	7.00	97.59	453.16	649	! Q20442 caenorhabditis elegans	sp_bacteria:Q32863	-	7.00	95.40	425.41	915	! Q32863 deinococcus himinis. f
sp_invertebrate:Q16242	+	7.00	97.57	452.90	651	! Q16242 caenorhabditis elegans	sp_bacteria:Q9LRUX2	-	7.00	95.40	425.32	916	! Q9lrux2 deinococcus radiodura
sp_bacteria:Q9Z8C4	+	7.00	97.55	452.64	653	! Q9z8c4 chlamydia pneumoniae (ch	sp_plant:Q9LMP6	+	7.00	95.36	424.89	921	! Q9lmp6 arabidopsis thaliana
sp_bacteria:Q9JSF0	+	7.00	97.55	452.64	653	! Q9jsf0 chlamydia pneumoniae (ch	sp_invertebrate:Q46354	+	7.00	95.36	424.81	922	! Q46354 caenorhabditis eleg
sp_virus:Q93130	-	7.00	97.47	451.63	661	! Q93130 coxsackievirus (cpv). d81	sp_invertebrate:Q9P2B2	+	7.00	95.34	424.64	924	! Q9p2b2 homo sapiens (human)
sp_plant:Q9M2R4	+	7.00	97.46	451.51	662	! Q9m2r4 arabidopsis thaliana (mc)	sp_invertebrate:Q57452	-	7.00	95.22	423.14	942	! Q57452 fundulus heteroclitus
sp_invertebrate:P81538	+	7.00	97.38	450.39	671	! P81538 drosophila ananassae	sp_invertebrate:Q9VCE7	-	7.00	95.19	422.97	944	! Q9vce7 drosophila melanoga
sp_invertebrate:Q9N917	+	7.00	97.38	450.39	671	! Q9n917 leishmania major. hyp	sp_virus:Q69139	+	7.00	95.19	422.81	946	! Q69139 human herpesvirus 4 (
sp_fungi:Q9YAF7	-	7.00	97.29	449.28	680	! Q9yaf7 exophiala dermatitidis (f	sp_plant:Q22913	+	7.00	95.13	422.07	955	! Q22913 arabidopsis thaliana
sp_invertebrate:Q9W531	+	7.00	97.14	447.37	696	! Q9w531 drosophila melanogaster	sp_human:Q60282	+	7.00	95.12	421.91	957	! Q60282 homo sapiens (human)
sp_bacteria:Q34003	+	7.00	97.11	446.89	700	! Q34003 rhodococcus sp. phaeoide	sp_invertebrate:Q18359	-	7.00	95.07	421.75	959	! Q18359 drosophila melanoga
sp_invertebrate:Q9VB25	+	7.00	97.11	446.89	700	! Q9vb25 drosophila melanogaster	sp_invertebrate:Q10953	+	7.00	95.07	421.34	964	! Q10953 caenorhabditis eleg
sp_organelle:Q9TLE9	+	7.00	97.08	446.54	703	! Q9tle9 nematode strumosa. nadh	sp_human:Q75179	+	7.00	95.05	421.02	968	! Q75179 homo sapiens (human)
sp_bacteria:Q9R2F3	-	7.00	97.08	446.54	703	! Q9rf3 deinococcus radiodurans	sp_virus:Q9J722	+	7.00	95.01	420.62	973	! Q9j722 parvovirus. prot
sp_invertebrate:Q16241	+	7.00	97.08	446.54	703	! Q16241 caenorhabditis elegans	sp_human:Q93387	+	7.00	94.97	420.07	980	! Q93387 homo sapiens (human)
sp_rodent:Q9WVM9	+	7.00	97.07	446.43	704	! Q9wvm9 mus musculus (mouse). ly	sp_invertebrate:Q76351	-	7.00	94.90	419.28	990	! Q76351 giardia lamblia (gi
sp_invertebrate:Q9N200	+	7.00	97.04	446.08	707	! Q9n2u0 caenorhabditis elegans	sp_invertebrate:Q9W2L7	-	7.00	94.87	418.90	995	! Q9w2l7 drosophila melanoga
sp_bacteria:Q9K502	+	7.00	97.02	445.73	710	! Q9k502 bacillus halodurans. cad	sp_bacteria:P95542	+	7.00	94.82	418.28	1003	! P95542 pseudomonas putida.
sp_bacteria:Q85906	-	7.00	97.02	445.73	710	! Q85906 sphingomonas aromaticiv	sp_human:Q43424	-	7.00	94.79	417.97	1007	! Q43424 homo sapiens (human)
sp_rodent:Q54729	-	7.00	97.00	445.50	712	! Q54729 rattus norvegicus (rat).	sp_rodent:Q61625	-	7.00	94.79	417.97	1007	! Q61625 mus musculus (mouse)
sp_invertebrate:Q9U0Z4	+	7.00	96.91	444.36	722	! Q9u0z4 leishmania major. hyp	sp_rodent:Q63226	-	7.00	94.79	417.90	1008	! Q63226 rattus norvegicus (r
sp_plant:Q9LMK3	+	7.00	96.88	443.91	726	! Q9lmk3 arabidopsis thaliana (mc)	sp_plant:Q9XK5	+	7.00	94.63	416.02	1033	! Q9xk5 chlamydomonas reinha
sp_vertebrate:Q9W7F9	-	7.00	96.87	443.79	727	! Q9w7f9 gallus gallus (chicken).	sp_rodent:Q35911	-	7.00	94.62	415.87	1035	! Q35911 mus musculus (mouse)
sp_phage:Q92XA5	+	7.00	96.85	443.57	729	! Q92xa5 bacteriophage phi-c31. g	sp_rodent:Q08684	-	7.00	94.61	415.80	1036	! Q08684 cricetus griseus (r
sp_bacteria:Q9KNQ3	+	7.00	96.81	443.01	734	! Q9knq3 vibrio cholerae. primosc	sp_rodent:Q35856	-	7.00	94.61	415.80	1036	! Q35856 rattus norvegicus (r
sp_rodent:Q9Q102	-	7.00	96.77	442.57	738	! Q9qq02 rattus norvegicus (rat).							



sp_rodent:O08959	-	7.00	94.61	415.80	1036	I_008959 rattus norvegicus (rat)
sp_invertebrate:Q9VZ30	+	7.00	94.59	415.50	1040	I_09vz30 drosophila melanogaster
sp_vertebrate:Q90840	+	7.00	94.59	415.50	1040	I_090840 gallus gallus (chicken)
sp_plant:Q91M01	-	7.00	94.58	415.35	1042	I_091m01 zea mays (maize). calca
sp_invertebrate:O45962	-	7.00	94.54	414.92	1048	I_045962 caenorhabditis elegans
sp_invertebrate:Q9VIV2	+	7.00	94.46	413.90	1062	I_09viv2 drosophila melanogaster
sp_bacteria:Q9JF62	+	7.00	94.41	413.40	1069	I_09jy62 neisseria meningitidis
sp_bacteria:Q9JF46	+	7.00	94.41	413.40	1069	I_09jy46 neisseria meningitidis
sp_human:Q9JF98	-	7.00	94.41	413.40	1069	I_09uf98 homo sapiens (human). h
sp_rodent:O54765	-	7.00	94.38	413.05	1074	I_054765 rattus norvegicus (rat)
sp_rodent:Q9Z280	-	7.00	94.38	413.05	1074	I_09z280 mus musculus (mouse). f
sp_invertebrate:Q19192	-	7.00	94.37	412.84	1077	I_019192 caenorhabditis elegans
sp_invertebrate:Q9VAM3	-	7.00	94.33	412.35	1084	I_09vam3 drosophila melanogaster
sp_human:Q73808	-	7.00	94.31	412.21	1086	I_075808 homo sapiens (human). s
sp_invertebrate:O44007	+	7.00	94.30	412.00	1089	I_044007 dictyostelium discoide
sp_plant:Q9SM80	-	7.00	94.26	411.58	1095	I_09sm80 oryza sativa (rice). zh
sp_rodent:Q93M80	+	7.00	94.26	411.58	1095	I_035867 rattus norvegicus (rat)
sp_rodent:Q9JL68	-	7.00	94.26	411.58	1095	I_09jlg8 mus musculus (mouse). s
sp_bacteria:O51934	+	7.00	94.21	410.96	1104	I_051934 thermotoga maritima. re
sp_invertebrate:Q9NED2	+	7.00	94.15	410.28	1114	I_09ned2 leishmania major. hyp
sp_human:Q9NQM5	-	7.00	94.14	410.14	1116	I_09nm5 homo sapiens (human). b
sp_fungi:Q12734	+	7.00	94.11	409.81	1121	I_012734 saccharomyces cerevisia
sp_invertebrate:Q9U2C0	+	7.00	94.09	409.61	1124	I_09u2c0 caenorhabditis elegans
sp_fungi:Q9P664	+	7.00	94.05	409.07	1132	I_09p664 neurospora crassa. rela
sp_fungi:Q12139	+	7.00	94.04	409.01	1133	I_012139 saccharomyces cerevisia
sp_human:Q13082	+	7.00	93.94	407.76	1152	I_013082 homo sapiens (human). m
sp_vertebrate:Q90783	+	7.00	93.82	406.40	1173	I_090783 gallus gallus (chicken)
sp_plant:Q65507	-	7.00	93.79	406.02	1179	I_065507 arabidopsis thaliana (bl
sp_invertebrate:Q94447	+	7.00	93.77	405.77	1183	I_094447 calliphora vicina (bl
sp_human:Q9JNG7	-	7.00	93.64	404.27	1207	I_09jng7 homo sapiens (human). t
sp_rodent:Q9QX70	+	7.00	93.63	404.15	1209	I_09qx70 rattus norvegicus (rat)
sp_human:Q60305	+	7.00	93.62	403.96	1212	I_060305 homo sapiens (human). k
sp_invertebrate:Q17240	+	7.00	93.59	403.66	1217	I_017240 bombyx mori (silk mot
sp_invertebrate:Q18270	+	7.00	93.58	403.60	1218	I_018270 caenorhabditis elegans
sp_invertebrate:Q77285	-	7.00	93.50	402.63	1234	I_077285 drosophila melanogaster
sp_invertebrate:Q9W0B8	-	7.00	93.50	402.63	1234	I_09w0b8 drosophila melanogaster
sp_bacteria:Q69947	-	7.00	93.48	402.39	1238	I_069947 streptomyces coelicolor
sp_invertebrate:Q10908	-	7.00	93.38	401.21	1257	I_010908 caenorhabditis elegans
sp_invertebrate:Q9V717	-	7.00	93.28	400.12	1277	I_09v717 drosophila melanogaster
sp_plant:O64496	-	7.00	92.95	396.30	1345	I_064496 arabidopsis thaliana (m
sp_vertebrate:Q00546	-	7.00	92.91	395.87	1353	I_000546 gallus gallus (chicken)
sp_invertebrate:Q9NEB8	+	7.00	92.90	395.71	1356	I_09neb8 leishmania major. hyp
sp_virus:Q89442	-	7.00	92.82	394.75	1374	I_089442 herpes simplex virus (h
sp_invertebrate:Q23371	-	7.00	92.65	392.87	1410	I_023371 caenorhabditis elegans
sp_invertebrate:Q9W3A4	+	7.00	92.62	392.52	1417	I_09w3a4 drosophila melanogaster
sp_bacteria:Q85862	-	7.00	92.59	392.11	1425	I_085862 sphingomonas aromatic
sp_bacteria:Q9KA72	-	7.00	92.55	391.71	1433	I_09ka72 bacillus halodurans. dr
sp_plant:Q9S7A0	-	7.00	92.52	391.35	1440	I_09s7a0 arabidopsis thaliana (m
sp_virus:Q91DV9	+	7.00	92.48	390.91	1449	I_091dv9 human immunodeficiency
sp_human:Q9UQP1	+	7.00	92.34	389.34	1481	I_09uqp1 homo sapiens (human). h
sp_invertebrate:Q9V4C8	+	7.00	92.26	388.47	1499	I_09v4c8 drosophila melanogaster
sp_invertebrate:Q21442	-	7.00	92.18	387.57	1518	I_021442 caenorhabditis elegans
sp_human:Q9UEH6	-	7.00	91.85	383.84	1600	I_09ueh6 homo sapiens (human). a
sp_invertebrate:Q9VYW1	+	7.00	91.66	381.80	1647	I_09vyw1 drosophila melanogaster
sp_human:Q12823	-	7.00	91.50	380.00	1690	I_012823 homo sapiens (human). a
sp_human:Q9UW66	-	7.00	91.50	380.00	1690	I_09uw66 homo sapiens (human). d
sp_human:Q9Y4L4	-	7.00	91.50	379.96	1691	I_09y4l4 homo sapiens (human). d
sp_plant:Q9S802	-	7.00	91.45	379.46	1703	I_09s802 arabidopsis thaliana (m
sp_human:Q9A4895	-	7.00	91.43	379.26	1708	I_09a4895 homo sapiens (human). k
sp_human:Q9UJP5	+	7.00	91.41	378.97	1715	I_09ujp5 homo sapiens (human). d
sp_human:Q13029	+	7.00	91.39	378.81	1719	I_013029 homo sapiens (human). r
sp_rodent:Q60592	-	7.00	91.34	378.20	1734	I_060592 mus musculus (mouse). r
sp_human:Q14675	-	7.00	91.30	377.76	1745	I_014675 homo sapiens (human). k
sp_human:Q9V6B8	-	7.00	91.19	376.66	1773	I_09v6b8 homo sapiens (human). s
sp_bacteria:Q9L8C8	-	7.00	90.99	374.40	1832	I_09l8c8 sorangium cellulosum. f
sp_bacteria:Q9KZ88	-	7.00	90.99	374.40	1832	I_09kz88 sorangium cellulosum. f
sp_invertebrate:Q9V9S7	+	7.00	90.95	373.95	1844	I_09v9s7 drosophila melanogaster
sp_invertebrate:Q9VWN8	-	7.00	90.94	373.87	1846	I_09vwn8 drosophila melanogaster
sp_invertebrate:Q9W4M7	-	7.00	90.72	371.50	1911	I_09w4m7 drosophila melanogaster
sp_human:Q43355	-	7.00	90.68	371.11	1922	I_043355 homo sapiens (human). f
sp_human:Q92508	-	7.00	90.32	367.23	2035	I_092508 homo sapiens (human). m
sp_human:Q75055	-	7.00	90.26	366.57	2055	I_075055 homo sapiens (human). m

sp_plant:Q9LGG0	+	7.00	90.25	366.47	2058	I_09lgg0 oryza sativa (rice)
sp_invertebrate:Q9VWE6	+	7.00	90.16	365.53	2087	I_09vwe6 drosophila melanog
sp_bacteria:Q45624	-	7.00	90.06	364.54	2118	I_045624 mycobacterium leprae
sp_rodent:Q9Z305	+	7.00	89.91	362.95	2169	I_09z305 cavia porcellus (gui
sp_human:Q9UHP0	-	7.00	89.91	362.88	2171	I_09uhp0 homo sapiens (human)
sp_invertebrate:Q9NGQ2	+	7.00	89.81	361.85	2205	I_09ngq2 dictyostelium disc
sp_invertebrate:Q9V515	+	7.00	89.80	361.76	2208	I_09v515 caenorhabditis ele
sp_human:Q93072	-	7.00	89.79	361.64	2212	I_093072 homo sapiens (human)
sp_human:Q9V5T2	-	7.00	89.70	360.71	2243	I_09v5t2 homo sapiens (human)
sp_human:Q9V5T2	-	7.00	89.68	360.51	2250	I_09v5t2 homo sapiens (human)
sp_human:Q9NYU8	-	7.00	89.65	360.18	2261	I_09nyu8 homo sapiens (human)
sp_human:Q9NYU4	-	7.00	89.63	360.04	2266	I_09nyu4 homo sapiens (human)
sp_human:Q9NYV1	-	7.00	89.61	359.83	2273	I_09nyv1 homo sapiens (human)
sp_human:Q9NYU6	-	7.00	89.58	359.51	2284	I_09nyu6 homo sapiens (human)
sp_human:Q9NYU9	-	7.00	89.54	359.11	2298	I_09nyu9 homo sapiens (human)
sp_invertebrate:Q9NG01	-	7.00	89.54	359.02	2301	I_09ng01 leishmania major.
sp_human:Q9NYU7	-	7.00	89.52	358.88	2306	I_09nyu7 homo sapiens (human)
sp_human:Q9NYV0	-	7.00	89.48	358.45	2321	I_09nyv0 homo sapiens (human)
sp_invertebrate:O16302	-	7.00	89.46	358.28	2327	I_016302 caenorhabditis ele
sp_human:Q9NYU5	-	7.00	89.45	358.14	2332	I_09nyu5 homo sapiens (human)
sp_human:Q9UHN9	-	7.00	89.33	356.88	2377	I_09uhn9 homo sapiens (human)
sp_human:Q9NS88	-	7.00	88.99	353.43	2506	I_09ns88 homo sapiens (human)
sp_invertebrate:Q21980	+	7.00	88.93	352.78	2531	I_021980 caenorhabditis ele
sp_human:Q9NY15	-	7.00	88.83	351.79	2570	I_09ny15 homo sapiens (human)
sp_plant:Q9XIR5	-	7.00	88.52	348.66	2698	I_09xir5 arabidopsis thaliana
sp_invertebrate:Q9XZB8	+	7.00	88.50	348.38	2710	I_09xzb8 plasmodium falcipa
sp_invertebrate:Q9NKP7	+	7.00	87.99	343.28	2936	I_09nkp7 leishmania major.
sp_rodent:Q9QXJ5	+	7.00	87.84	341.80	3006	I_09qxj5 mus musculus (mouse)
sp_bacteria:Q9S582	-	7.00	87.79	341.32	3029	I_055582 synecocystis sp. (s
sp_virus:Q89459	-	7.00	87.59	339.42	3122	I_089459 herpes simplex virus
sp_vertebrate:Q9W6V0	-	7.00	86.70	330.73	3595	I_09w6v0 gallus gallus (chick
sp_invertebrate:P91904	-	7.00	86.51	328.91	3704	I_091904 caenorhabditis ele
sp_human:Q9Y6V2	-	7.00	86.18	325.82	3899	I_09y6v2 homo sapiens (human)
sp_human:Q9UQ04	-	7.00	86.17	325.69	3908	I_09uq04 homo sapiens (human)
sp_human:Q99996	-	7.00	86.16	325.64	3911	I_099996 homo sapiens (human)
sp_rodent:Q61291	-	7.00	85.21	316.76	4345	I_061291 mus musculus (mouse)
sp_rodent:Q9QXR6	+	7.00	84.81	313.20	4833	I_09qxr6 mus musculus (mouse)
sp_rodent:Q9JRT1	+	7.00	84.75	312.65	4880	I_09jrt1 rattus norvegicus (r
sp_rodent:Q9QXV7	+	7.00	84.55	310.82	5038	I_09qxv7 mus musculus (mouse)
sp_bacteria:Q9S2545	-	7.00	84.52	310.57	5060	I_052545 amycolactopsis medit
sp_bacteria:Q9S2789	-	7.00	84.51	310.47	5069	I_052789 amycolactopsis medite
sp_rodent:Q9JKS6	+	7.00	84.49	310.29	5085	I_09jks6 rattus norvegicus (r
sp_invertebrate:Q9NI64	+	7.00	84.20	307.70	5322	I_09ni64 drosophila melanog
sp_bacteria:Q9R9J0	+	7.00	84.15	307.20	5369	I_09r9j0 bacillus subtilis. i
sp_bacteria:Q85168	-	7.00	80.60	275.72	9376	I_085168 pseudomonas syringae
sp_bacteria:Q9L4X3	+	7.00	80.53	275.16	9477	I_09l4x3 streptomyces noursei

seq\_name: sp\_bacteria:Q9JVT1

seq\_documentation\_block:

ID\_Q9JVT1 PRELIMINARY;

AC Q9JVT1; PRT: 499 AA.

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DE 01-OCT-2000 (TRENBLrel. 15, Last annotation update)

GN PUTATIVE PERIPLASMIC SERINE PROTEASE (EC 3.4.21.1).

NCBI\_taxid:65699;

OC Neisseria meningitidis (serogroup A).

OS Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

RX STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;

RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,

RA Davies R.M., Davis P., Basham D., Brown D., Chillingworth T.,

RA Jørgensen K., Leatham S., Moule S., Mungall K., Quail M.A.,

RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,

RA Whitehead S., Spratt B.G., Barrell B.G.;

RT "Complete DNA sequence of a serogroup A strain of Neisseria

meningitidis 22491.";

RL Nature 404:502-506(2000).  
DR ENBL; AL162754; CAB83996.1; -.  
KW Protease: Hydrolase.  
SQ SEQUENCE 499 AA; 52563 MW; 73E979F0A7F5B969 CRC64;

alignment\_scores:  
Quality: 127.00 Length: 127  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Q9JVT1 ..

Align seg 1/1 to: Q9JVT1 from: 1 to: 499

583 TTGAACCGGGCGAATGGTCTGCTCCATCGCGCGCCCTTCGGCTTTGA 632  
|||||  
195 LeuLysProGlyGlyTrpValAlaAlaAlaGlyAlaProPheGlyPheAs 211  
633 CAACAGCGTGACCGCGCGCATGCTGCTCCGCCAAAGCAGAAAGCCTGCCCA 682  
|||||  
211 pAsnSerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProA 228  
683 ACGAAAGCTACACACCCCTTCATCCAAACGACGCTTGCATCAATCCGGGC 732  
|||||  
228 snGluSerTyThrProPheIleGlnThrAspValAlaIleAsnProGly 244  
733 AATTCCGCGCGCGCTGTTCACCTTAAAGACACAGTCTCGGCATCAA 782  
|||||  
245 AsnSerGlyGlyProLeuPheAsnLeuLysGlyGlnValValGlyIleAs 261  
783 TTCGCAATATACACCGCGCGGATTCATGGGCATCTCCTTTGCCA 832  
|||||  
261 nSerGlnIleTySerArgSerGlyGlyPheMetGlyIleSerPheAlaI 278  
833 TCCCGATTACGCTTCCCATGAATGTCGCGAAGACAGCTGAAACACCGGC 882  
|||||  
278 leProIleAspValAlaMetAsnValAlaGluGlnLeuLysAsnThrGly 294  
883 AAAGTCCACGCGGACAACTGGCGCTGATTATTACGGAAGTATCTACGG 932  
|||||  
295 LysValGlnArgGlyGlnLeuGlyValIleIleGlnGluValSerTyGly 311  
933 TTTGGCACAGTGTTCGGTCTCGGATAAAGCC 963  
|||||  
311 yLeuAlaGlnSerPheGlyLeuAspLysAla 321

seq\_name: sp\_bacteria:Q44476

seq\_documentation\_block:  
ID Q44476 PRELIMINARY; PRT: 473 AA.  
AC Q44476;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE MUCD.  
OS Azotobacter vinelandii.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Azotobacter.  
OX NCBI\_TaxID=354;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=UW 136;  
RX MEDLINE=96178940; PubMed=8606151;  
RA Martinez-Salazar J.M., Moreno S., Najera R., Boucher J.C., Espin G.,  
RA Soberon-Chavez G., Deretic V.;  
RT "Characterization of the genes coding for the putative sigma factor  
RT Algu and its regulators MucA, MucB, MucC, and MucD in Azotobacter  
RT vinelandii and evaluation of their roles in alginate biosynthesis.";  
RL J. Bacteriol. 178:1800-1808(1996).  
DR ENBL; U30799; AAB01513.1; -.  
DR INTERPRO; IPR001254; -.  
DR

DR INTERPRO: IPR001478; -.  
DR PFAM; PF00089; trypsin; 2.  
DR PFAM; PF00595; PDZ; 2.  
SQ SEQUENCE 473 AA; 50282 MW; C97B357D897738AB CRC64;

alignment\_scores:  
Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Q44476 ..

Align seg 1/1 to: Q44476 from: 1 to: 473

697 CCCTTCATCCAAACGACGTTGCCATCAATCGGGCAATTCGGCGGCC 746  
|||||  
204 ProPheIleGlnThrAspValAlaIleAsnProGlyAsnSerGlyGlyPr 220  
747 GCTGTTTC 753  
|||||  
220 oLeuPhe 222

seq\_name: sp\_bacteria:Q57155

seq\_documentation\_block:

ID Q57155 PRELIMINARY; PRT: 474 AA.  
AC Q57155;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)  
DE MUCD.  
GN MUCD.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=93391358; PubMed=8378309;  
RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,  
RA Deretic V.;  
RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa  
RT infecting cystic fibrosis patients.";  
RL Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=95286510; PubMed=7768826;  
RA Yu H., Schurr M.J., Deretic V.;  
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas  
RT aeruginosa Algu: E. coli rpoE restores mucoidy and reduces sensitivity  
RT to reactive oxygen intermediates in algu mutants of P. aeruginosa.";  
RL J. Bacteriol. 177:3259-3268(1995).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PA01;  
RX MEDLINE=96134987; PubMed=8550474;  
RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,  
RA Deretic V.;  
RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas  
RT aeruginosa in cystic fibrosis encode homologs of the serine protease  
RT HtrA.";  
RL J. Bacteriol. 178:511-523(1996).  
DR ENBL; U49151; AAC43718.1; -.  
DR EMBL; U32853; AAC43676.1; -.  
DR INTERPRO; IPR001254; -.  
DR INTERPRO; IPR001478; -.  
DR INTERPRO; IPR001940; -.  
DR PFAM; PF00089; trypsin; 1.  
DR PFAM; PF00595; PDZ; 2.  
DR

DR PRINTS: PRO0834; PROTEASES2C.  
SQ SEQUENCE 474 AA; 50321 MW; 8AB5D2A89867BEBC CRC64;

alignment\_scores:  
Quality: 19.00 Length: 19  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

-alignment\_block:

US-09-388-090-3 x Q57155 ..

Align seg 1/1 to: Q57155 from: 1 to: 474

637 AGCGTGACCGCGCATTCGGTCCGCAAGGAGAGCCCTGCCCAACGA 686

|||||

184 SerValThrAlaGlyIleValSerAlaLysGlyArgSerLeuProAsnG1 200

687 AACCTAC 693

|||||

200 userTyf 202

seq\_name: sp\_bacteria:Q9KJN6

seq\_documentation\_block:

ID Q9KJN6 PRELIMINARY; PRT; 500 AA.

AC Q9KJN6;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)

DE PUTATIVE SERINE PROTEASE DO-LIKE PRECURSOR.

GN HTRA.

OS Myxococcus xanthus.

OC Bacteria; Proteobacteria; delta subdivision; Myxobacteria;

OC Myxococcales; Cystobacterineae; Myxococcaceae; Myxococcus.

OX NCBI\_TaxID=34;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=D22;

RX MEDLINE=20032045; PubMed=10564511;

RA Cho K., Zusman D.R.;

RT "Sporulation timing in Myxococcus xanthus is controlled by the espAB

RT locus";

RL Mol. Microbiol. 34:714-725(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=D22;

RA Cho K., Zusman D.R.;

RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF163841; AAF87931.1; ..

KW Protease.

SQ SEQUENCE 500 AA; 50498 MW; 012DCAE1F38B9CDD CRC64;

alignment\_scores:

Quality: 13.00 Length: 13  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Q9KJN6 ..

Align seg 1/1 to: Q9KJN6 from: 1 to: 500

718 GCCATCAATCCGGCGCAATTCGGCGGCCCGCTGTTCAC 756

|||||

245 AlaIleAsnProGlyAsnSerGlyGlyProLeuPheAsn 257

seq\_name: sp\_bacteria:Q35021

seq\_documentation\_block:

ID Q35021

AC Q35021; PRELIMINARY; PRT; 224 AA.

DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)  
DE YVTB PROTEIN.  
GN Bacillus subtilis.  
OS Bacteria; Firmicutes; Bacillus/Clostridium group;  
OC Bacillus/Staphylococcus group; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RX MEDLINE=98044033; PubMed=9384377;  
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,  
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,  
RA Borilliet S., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,  
RA Brullis R., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,  
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,  
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,  
RA Frits C., Fujita M., Fujita Y., Fuma S., Galizzi A., Foulger D.,  
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Galleron N.,  
RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,  
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,  
RA Joris B., Karamata D., Kasahara Y., Kleerr-Blanchard M., Klein C.,  
RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,  
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,  
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,  
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,  
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,  
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,  
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,  
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,  
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,  
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,  
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,  
RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Tognoni A.,  
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,  
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,  
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,  
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;  
RT "The complete genome sequence of the gram-positive bacterium Bacillus  
RT subtilis";  
RL Nature 390:249-256(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=168;  
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;  
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98015415; PubMed=9353931;  
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;  
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289  
RT degrees) in Bacillus subtilis";  
RL Microbiology 143:3305-3308(1997).  
DR EMBL; Z99120; CAB15289.1; ..  
DR EMBL; Z93941; CAB07969.1; ..  
DR HSSP; P09331; IAGJ.  
DR INTERPRO; IPR001254; ..  
DR INTERPRO; IPR001478; ..  
DR PFAM; PF00089; trypsin; 1.  
DR PFAM; PF00595; PDZ; 1.  
SQ SEQUENCE 224 AA; 23699 MW; 34A7CF0D79C7FB59 CRC64;

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x Q35021 ..



Align seg 1/1 to: 035021 from: 1 to: 224

718 GCCATCAATCGGGCAATTCCGGCGCGCGCTG 750  
|||||  
58 AlaIleAsnProGlyAsnSerGlyGlyProLeu 68

seq\_name: sp\_bacteria:p72780

seq\_documentation\_block:

ID P72780 PRELIMINARY; PRT; 394 AA.

DT 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE PUTATIVE SERINE PROTEASE.

GN HHOA.

OS Synechocystis sp. (strain PCC 6803)

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI\_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-97061201; PubMed-8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirose M., Sugiura M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.;"

RL DNA Res. 3:109-136(1996).

DR EMBL; D90900; BAA16795.1; -.

DR MEROPS; S01.274; -.

DR INTERPRO; IPR001254; -.

DR INTERPRO; IPR001478; -.

DR INTERPRO; IPR001940; -.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00595; PDZ; 1.

DR PRINTS; PR00834; PROTEASES2C.

KW Protease.

SQ SEQUENCE 394 AA; 41336 MW; AD59D94811B8F57B CRC64;

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x P72780 ..

Align seg 1/1 to: P72780 from: 1 to: 394

718 GCCATCAATCGGGCAATTCCGGCGCGCGCTG 750  
|||||  
231 AlaIleAsnProGlyAsnSerGlyGlyProLeu 241

seq\_name: sp\_bacteria:006670

seq\_documentation\_block:

ID 006670 PRELIMINARY; PRT; 397 AA.

AC 006670;

DT 01-JUL-1997 (TREMBLrel. 04, Created)

DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE PUTATIVE SERINE PROTEASE.

GN SPHRA.

OS Streptococcus pneumoniae.

OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

OC Streptococcus.

OX NCBI\_TaxID=1313;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-R801;

RA Gasc A.A.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF000658; AAC45334.1; -.

DR MEROPS; S01.273; -.

DR INTERPRO; IPR001254; -.

DR INTERPRO; IPR001478; -.

DR INTERPRO; IPR001940; -.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00595; PDZ; 1.

DR PRINTS; PR00834; PROTEASES2C.

KW Protease.

SQ SEQUENCE 397 AA; 42287 MW; 755C13398AB8EE15 CRC64;

alignment\_scores:

Quality: 11.00 Length: 11  
Ratio: 1.000 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000

alignment\_block:

US-09-388-090-3 x 006670 ..

Align seg 1/1 to: 006670 from: 1 to: 397

718 GCCATCAATCGGGCAATTCCGGCGCGCGCTG 750

|||||

232 AlaIleAsnProGlyAsnSerGlyGlyProLeu 242

seq\_name: sp\_bacteria:083752

seq\_documentation\_block:

ID 083752 PRELIMINARY; PRT; 398 AA.

AC 083752;

DT 01-NOV-1998 (TREMBLrel. 08, Created)

DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE PERIPLASMIC SERINE PROTEASE DO (HTRA-1).

GN TP0773.

OS Treponema pallidum.

OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.

OX NCBI\_TaxID=160;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-NICHOLS;

RX MEDLINE-98332770; PubMed-9665876;

RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,

RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,

RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,

RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,

RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,

RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,

RA Venter J.C.;

RT "Complete genome sequence of Treponema pallidum, the syphilis

RT spirochete.;"

RL Science 281:375-388(1998).

DR EMBL; AE001248; AAC65740.1; -.

DR HSSP; P41140; 2SFA.

DR MEROPS; S01.274; -.

DR TIGR; TP0773; -.

DR INTERPRO; IPR001254; -.

DR INTERPRO; IPR001314; -.

DR INTERPRO; IPR001478; -.

DR INTERPRO; IPR001940; -.

DR PFAM; PF00089; trypsin; 1.

DR PFAM; PF00595; PDZ; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR PRINTS; PR00834; PROTEASES2C.

KW Protease.

SQ SEQUENCE 398 AA; 42186 MW; 8715EA05F454258D CRC64;

alignment\_scores:  
 Quality: 11.00 Length: 11  
 Ratio: 1.000 Gaps: 0  
 Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x 083752 ..

Align seg 1/1 to: 083752 from: 1 to: 398

718 GCCATCAATCCGGCAATTCCGGCGCGCGCTG 750

|||||  
 223 AlaIleAsnProGlyAsnSerGlyGlyProLeu 233

seq\_name: sp\_plant:004674

seq\_documentation\_block:

ID 004674 PRELIMINARY; PRT; 398 AA.

AC 004674;

DT 01-JUL-1997 (TReMBLrel. 04, Created)

DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE HTRA-LIKE PROTEIN.

GN HTRA.

OS Haematococcus pluvialis.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Haematococaceae; Haematococcus.

OX NCBI\_TaxID=44745;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97324581; PubMed=9180686;

RA Hershkovits G., Dubinsky Z., Katcoff D.J.;

RT "A novel homologue of the prokaryotic htrA gene is differentially

expressed in the alga Haematococcus pluvialis following stress.";

RL Mol. Gen. Genet. 254:345-350(1997).

DR EMBL: U58980; AAB61311.1;

DR MEROPS: S01.279;

DR MENDEL: 23986; Haep1:3069; 23986.

DR INTERPRO: IPR001254;

DR INTERPRO: IPR001478;

DR INTERPRO: IPR001940;

DR PFAM: PF00089; trypsin; 1.

DR PFAM: PF00595; PDZ; 1.

DR PRINTS: PR00834; PROTEASES2C.

SQ SEQUENCE 398 AA; 41469 MW; D4D404273FB4309A CRC64;

## alignment\_scores:

Quality: 11.00 Length: 11

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x 004674 ..

Align seg 1/1 to: 004674 from: 1 to: 398

718 GCCATCAATCCGGCAATTCCGGCGCGCGCTG 750

|||||  
 237 AlaIleAsnProGlyAsnSerGlyGlyProLeu 247

seq\_name: sp\_bacteria:P73940

seq\_documentation\_block:

ID P73940 PRELIMINARY; PRT; 416 AA.

AC P73940;

DT 01-FEB-1997 (TReMBLrel. 02, Created)

DT 01-FEB-1997 (TReMBLrel. 02, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE PROTEASE HHOB.

GN HHOB.

OS Synecocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.

OX NCBI\_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hiroseawa M., Suglura M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimpou S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synecocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";

RL DNA Res. 3:109-136(1996).

DR EMBL: D90911; BAA18006.1;

DR MEROPS: S01.275;

DR INTERPRO: IPR001254;

DR INTERPRO: IPR001478;

DR INTERPRO: IPR001940;

DR PFAM: PF00089; trypsin; 1.

DR PFAM: PF00595; PDZ; 1.

DR PRINTS: PR00834; PROTEASES2C.

KW Protease.

SQ SEQUENCE 416 AA; 43196 MW; DD32BD997CC6B25B CRC64;

## alignment\_scores:

Quality: 11.00 Length: 11

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

## alignment\_block:

US-09-388-090-3 x P73940 ..

Align seg 1/1 to: P73940 from: 1 to: 416

718 GCCATCAATCCGGCAATTCCGGCGCGCGCTG 750

|||||

252 AlaIleAsnProGlyAsnSerGlyGlyProLeu 262

seq\_name: sp\_plant:Q9LK85

seq\_documentation\_block:

ID Q9LK85 PRELIMINARY; PRT; 439 AA.

AC Q9LK85;

DT 01-OCT-2000 (TReMBLrel. 15, Created)

DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TReMBLrel. 15, Last annotation update)

DE DEGP PROTEASE PRECURSOR.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;

OC Brassicales; Brassicaceae; Arabidopsids.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Nakamura Y.;

RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.

RT Sequence features of the regions of 4,251,695 bp covered by ninety P1,

RT TAC and BAC clones";

RL DNA Res. 7:217-221(2000).

DR EMBL: AP000371; BAB02539.1;

DR EMBL: AP001302; BAB02539.1; JOINED.

KW Protease.

SQ SEQUENCE 439 AA; 46673 MW; 05EB437DCE71A251 CRC64;

## alignment\_scores:

```

Quality: 11.00      Length: 11
Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x Q9LK85 ..
Align seg 1/1 to: Q9LK85 from: 1 to: 439

718 GCCATCAATCCGGCAATTCGGCGCCGCGCTG 750
|||||
276 AlaIleAsnProGlyAsnSerGlyGlyProLeu 286

seq_name: sp_bacteria:Q9RTK4

seq_documentation_block:
ID Q9RTK4 PRELIMINARY; PRT; 441 AA.
AC Q9RTK4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE PERPLASMIC SERINE PROTEASE DO, PUTATIVE..
GN DRI1756.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eilen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the Radioresistant Bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
DR EMBL; AE002017; AAFI1312.1; -.
DR HSP; P00778; ITAL.
DR TIGR; DR1756; -.
DR INTERPRO; IPR001254; -.
DR INTERPRO; IPR001478; -.
DR INTERPRO; IPR001940; -.
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
KW Protease.
SQ SEQUENCE 441 AA; 45145 MW; D9E475AD081E0190 CRC64;

alignment_scores:
Quality: 11.00      Length: 11
Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x Q9RTK4 ..
Align seg 1/1 to: Q9RTK4 from: 1 to: 441

718 GCCATCAATCCGGCAATTCGGCGCCGCGCTG 750
|||||
256 AlaIleAsnProGlyAsnSerGlyGlyProLeu 266

seq_name: sp_plant:Q9LU10

seq_documentation_block:
ID Q9LU10 PRELIMINARY; PRT; 448 AA.
AC Q9LU10;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)

```

```

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, TAC CLONE:K13H13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB024023; BAA98101.1; -.
SQ SEQUENCE 448 AA; 47492 MW; A986FC1387670AFF CRC64;

```

```

alignment_scores:
Quality: 11.00      Length: 11
Ratio: 1.000        Gaps: 0
Percent Similarity: 100.000  Percent Identity: 100.000

```

```

alignment_block:
US-09-388-090-3 x Q9LU10 ..

```

```

Align seg 1/1 to: Q9LU10 from: 1 to: 448

```

```

718 GCCATCAATCCGGCAATTCGGCGCCGCGCTG 750
|||||
286 AlaIleAsnProGlyAsnSerGlyGlyProLeu 296

```

```

seq_name: sp_bacteria:O34358

```

```

seq_documentation_block:
ID O34358 PRELIMINARY; PRT; 449 AA.
AC O34358;

```

```

DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DE YKDA.
GN YKDA OR HTRA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Halech J., Hartwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,

```

```
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zunstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ002571; CAA05570.1; -
DR EMBL; Z99110; CAB13147.1; -
DR MEROPS; S01.273; -
DR INTERPRO; IPR001254; -
DR INTERPRO; IPR001478; -
DR INTERPRO; IPR001940; -
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
KW Protease.
SQ SEQUENCE 449 AA; 47713 MW; E12B07A9018EE414 CRC64;

alignment_scores:
  Quality: 11.00 Length: 11
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x 034358 ..
Align seg 1/1 to: 034358 from: 1 to: 449

718 GCCATCAATCCGGCAATTCGGCGCGCCGCTG 750
|||||
284 AlaIleAsnProGlyAsnSerGlyGlyProLeu 294

seq_name: sp_bacteria:p73354

seq_documentation_block:
ID_ P73354 PRELIMINARY; PRT; 452 AA.
AC P73354;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE SERINE PROTEASE HTRA.
GN HTRA.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_taxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE-97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL; D90905; BAA17385.1; -
DR MEROPS; S01.273; -
```

```
DR INTERPRO; IPR001254; -
DR INTERPRO; IPR001478; -
DR PFAM; PF00089; trypsin; 1.
DR PFAM; PF00595; PDZ; 1.
KW Protease.
SQ SEQUENCE 452 AA; 47656 MW; AED4AFB1AC472361 CRC64;

alignment_scores:
  Quality: 11.00 Length: 11
  Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-388-090-3 x P73354 ..
Align seg 1/1 to: P73354 from: 1 to: 452

718 GCCATCAATCCGGCAATTCGGCGCGCCGCTG 750
|||||
290 AlaIleAsnProGlyAsnSerGlyGlyProLeu 300
```

---

